With regard to the new claims 5-20, which relate to the δ′ subunit of polymerase III holoenzyme, it is well known to one skilled in the art that proteins homologous to this subunit of polymerase III holoenzyme are contained in organisms other than *E. coli*, as shown in the Declaration of Michael O'Donnell under 37 CFR § 1.132 submitted in parent U.S. Patent Application Serial No. 08/279,058 on December 17, 1996 ("O'Donnell Declaration") (copy submitted herewith).

Those skilled in the art recognize the δ′ subunit from *E. coli* has sequence homology to accessory protein complexes of various other organisms (O'Donnell Declaration ¶ 13). For example, in O'Donnell et al., "Homology in Accessory Proteins of Replicative Polymerases - *E. coli* to Humans," Nucleic Acids Research 21(1):1-3 (1993) ("O'Donnell")(copy attached hereto as Exhibit 1), a comparison of amino acid sequences shows the homology between proteins of replicative polymerases of *E. coli*, humans, and phage T4 (O'Donnell Declaration ¶ 13). In Carter, et al., "Identification, Isolation, and Characterization of the Structural Gene Encoding the δ′ Subunit of *Escherichia coli* DNA Polymerase III Holoenzyme," J. of Bacteriology, 175(12):3812-22 (1993), Figure 5 diagrams the homology of the δ′ amino acid sequence to other replication proteins (O'Donnell Declaration ¶ 13). Comparison of the δ′ amino acid sequence revealed similarity to the A1(replication factor C) complex of HeLa cells and to the gene 44 protein (gp44) of bacteriophage T4 (O'Donnell Declaration ¶ 13). In addition, amino acid sequence similarity was found to the gene product of *B. subtilis* (O'Donnell Declaration ¶ 13). Further, the structural homology of the δ′ subunit to other replication proteins has been proven to be true (O'Donnell Declaration ¶ 13). For example, the genome project of *Haemophilus influenze* showed homologues to all 10 subunits of *E. coli* DNA polymerase III holoenzyme, including δ, δ′, χ, Ψ, and θ (O'Donnell Declaration ¶ 13). Currently, the GenBank now also shows homologues to the δ′ subunit of *E. coli* from a large variety of organisms, including the following: Procaryotes:

Escherichia coli, Haemophilus influenze, Micrococcus luteus, Pseudomonas aeruginosa, Bacillus subtilis, Caulobacter crescentus; Archaebacteria: *Thermus thermophilis* (extreme thermophile); Eukaryotes: *Drosophila melanogaster* (fly, insect), *Caenorhabditis elegans* (namatode, worm), *Gallus gallus* (dog), *Homo sapien* (man), *Saccharomyces cerevisiae* (yeast), and *Saccharomyces pombe* (yeast) (O'Donnell Declaration ¶ 13).

Further, the sequence of the human homologues to δ', and indeed the other δ' homologues, are sufficiently homologous to the δ' subunit of *E. coli* to provide for identifying and obtaining the corresponding δ' (holA) gene from these organisms using the gene encoding the δ' subunit of *E. coli* in the following ways: (1) use of the *E. coli* holA gene, or fragments of the *E. coli* gene, as a probe in a Southern analysis of whole cell DNA from another organisms to identify the corresponding δ' homologue; (2) use of holA, or its fragments, as a probe to screen cDNA plasmid libraries of other organisms; (3) use of the holA gene sequence to synthesize oligonucleotide primers for PCR to amplify the corresponding δ' homologue from total genomic DNA from other organisms; and (4) use of the holA gene sequence to identify the δ' homologue from a genome sequencing project of other organisms by sequence comparison to the *E. coli* holA gene (O'Donnell Declaration ¶ 14).

The present application fully discusses the isolation and sequencing of the δ' subunit and its encoding genes for the polymerase III holoenzyme. In view of the disclosure of these experimental procedures, and the known structural and functional homology of the δ' subunit proteins from various sources such as yeast RFC subunits, human RFC subunits, bacteriophage T4 gp44 subunits, and DnaH subunits of *B. subtilis*, it would not require an undue amount of experimentation for one skilled in the art to isolate and sequence the claimed δ' protein (and its encoding gene) from sources other than *E. coli*.

With regard to the new claims 21-75, which relate to the protein subunits other than δ', it is well known to one skilled in the art that proteins in other organisms have functional and structural homology to the subunits of *E. coli* (O'Donnell Declaration ¶¶ 10-16).

As discussed above, in O'Donnell a comparison of amino acid sequences shows the homology between proteins of replicative polymerases of *E. coli*, humans, and phage T4. Further, in Sanders et al., "Rules Governing the Efficiency and Polarity of Loading a Tracking Clamp Protein Onto DNA: Determinants of Enhancement in Bacteriophage T4 Late Transcription," The EMBO Journal 14(16):3966-76 (1995)("Sanders") (copy attached hereto as Exhibit 2), the common elements of structure and function of replicative DNA polymerases of eukaryotes, prokaryotes, and certain viruses are discussed (O'Donnell Declaration ¶ 12). It is disclosed that the replicative DNA polymerases of all of these sources are composed of a core enzyme and a set of accessory proteins (O'Donnell Declaration ¶ 12). Further, Stillman, "Smart Machines at the DNA Replication Fork," Cell 78:725-28 (1994)("Stillman") (copy attached hereto as Exhibit 3) discusses the functional similarity of proteins from *E. coli*, humans, and phage T4 that cause replication (O'Donnell Declaration ¶ 12). Specifically, these exhibits show that *E. coli* contains an accessory complex called γ complex which contains the subunits γ, δ, δ', Ψ, and χ (O'Donnell Declaration ¶ 12). Further, these exhibits show that homologous proteins to the γ complex are also present in eukaryotic (containing RFC complex), phage T4 (containing g44 complex), and human (containing RFC complex) organisms (O'Donnell Declaration ¶ 12).

As further discussed in these references, all cellular replicases known to date utilize a DNA sliding clamp that encircles DNA and acts as a mobile tether to hold continuously the DNA polymerase to DNA as it moves along it. The clamp requires an accessory protein to assemble the sliding clamp around the DNA and, therefore, a clamp loader

that couples ATP hydrolysis to assemble the clamp onto DNA for use by the polymerase is required.

As disclosed in O'Donnell, Stillman, and Sanders, the clamp loader complex of *E. coli* is the γ complex, the clamp loader complex in human and eukaryotic organisms is the RFC complex, and the clamp loader complex in phage T4 is the g44 protein complex. These clamp loader complexes all contain 5 subunits, and are functionally similar in utilizing ATP to transfer their respective clamp onto DNA. Further, the subunits of the clamp loaders are homologous in their amino acid sequence from *E. coli* to humans, including yeast and bacteriophage T4 (See O'Donnell).

As also discussed in O'Donnell, the clamp of *E. coli* is called the β subunit. The crystal structure of β shows it is a ring shaped protein with 6 globular domains that encircle DNA. The functionally homologous clamp from yeast and humans is called proliferating cell nuclear antigen ("PCNA") and the functionally homologous clamp from bacteriophage T4 is called gp45. The structural homology of β to PCNA from yeast and humans and the phage T4 gene 45 protein was pointed out in Kong et al., "Three-Dimensional Structure of the β Subunit of E. coli DNA Polymerase III Holoenzyme: A Sliding DNA Clamp," Cell, 69:425-437 (1992)("Kong")(copy attached hereto as Exhibit 4). Further, the structural similarity was proven with the crystal structure of yeast PCNA (Krishna et al. "Crystal Structure of the Eukaryotic DNA Polymerase Processivity Factor PCNA", Cell 79:1233-43 (1994) ("Krishna") (copy attached hereto as Exhibit 5). Like β, PCNA is ring shaped with six domains and these domains have the exact same chain topology folding pattern as the *E. coli* β subunit clamp. Id.. The crystal structure of the bacteriophage T4 clamp gp45 shows it too is a 6 domain ring having the same chain fold as the clamps of *E. coli*, yeast, and humans. Examination of the GeneBank sequence information shows that all identified bacterial β subunit genes (the dnaN gene) are homologous to the β subunit of *E. coli* (Fig. 2 in Kelman, et. al., "Structural and Functional Similarities of Prokaryotic and

Eukaryotic DNA Polymerase Sliding Clamps," Nuc. Acids Res. 23:3613-20 (1995)("Kelman") (copy attached hereto as Exhibit 6). These organisms are: *Bacillus subtilis, Micrococcus luteus, Streptomyces coelicolor, Pseudomonas putida, Serratic marcenscens, Salmonella typhimurium, Proteus mirabilis, Mycoplasma capricolum and Actinobacillus pleuropneumoniae*. To carry the β to PCNA homology further, several eukaryotes have homologues of the clamp including: human, mouse, rat, frog (*Xenopus laevis*), fly (*Drosophila melanogaster*), *Catharanthus roseus*, carrot (*Daucus carrota*), *Glycine max, Oriza sativa, Saccharomyces pombe, Saccharomyces cerevisiae, and Autographa californica* (Fig. 3 in Kelman). Thus, the accessory proteins (i.e. the clamp and clamp loader) of *E. coli*, bacteriophage T4, yeast, and humans are known by those of ordinary skill in the art to have functional similarity.

Further, the functional similarity between the complexes of various organisms was demonstrated where the phage T4 proteins (clamp, clamp loader, and polymerase) were able to substitute for the human homologues in an *in vitro* assay of the SV40 replication system (Tsurimoto, et al., "Functions of Replication Factor C and Proliferating Cell Nuclear Antigen: Functional Similarity of DNA Polymerase Accessory Proteins From Human Cells and Bacteriophage T4," PNAS, 87:1023-1027 (1990)("Tsurimoto") (copy attached hereto as Exhibit 7) (O'Donnell Declaration ¶ 11).

Further, the sequence homology of the δ' subunit and the γ subunit in *E. coli* is disclosed in Dong et al., "DNA polymerase III accessory proteins, I. holA and holB encoding δ and δ'" J. Biol. Chem., 268:11758-65 (1993)("Dong") (copy attached hereto as Exhibit 8). This homology in two subunits of polymerase III holoenzyme suggests that these were partners rooted in a common ancestor gene which duplicated and then endured slow successive changes. Thus, it would be known by those of ordinary skill in the art that this homology in particular would be carried out in other organisms. Indeed, the homology of the δ' and γ subunits extends to all five

subunits of the RFC clamp loading complex of yeast and of humans (O'Donnell). They all have homology to the δ' and γ subunits of the E. coli γ complex.(Id.). Yeast and humans are the two most highly advanced eukaroyotic replication systems. The fact that yeast is a lower eukaryote and human is a higher eukaryote and that the two systems span a large evolutionary distance makes it likely that everything in between also will be homologous.

Accordingly, those of ordinary skill in the art would have been able to isolate the claimed protein subunits (and their encoding genes) from sources other than *E. coli*.

The rejection of claim 1 under 35 USC § 102(b) as anticipated by Takase et al., "Genes Encoding Two Lipoproteins in the *leuS-dac*A Region of the *Escherichia coli* Chromosome," J. of Bacteriology 169(12):5692-99 (1987) ("Takase") is respectfully traversed.

Takase relates to the coding of two lipoproteins by two genes, *rlpA* and *rlpB*, located in the *leuS-dac*A region on the *Escherichia coli* chromosome (O'Donnell Declaration ¶ 17). The *rlpA* gene encodes for a lipoprotein having molecular weight of 36K (O'Donnell Declaration ¶ 17). Figure 6 of the reference details the sequence of the 36K lipoprotein gene *rlpA* and its 5'- and 3'- flanking regions and the amino acid sequences deduced from the nucleotide sequence (O'Donnell Declaration ¶ 17). The position of the PTO is that this sequence matches that of the sequence encoding the claimed δ subunit. Applicants respectfully disagree. Takase also discloses a *rlpB* gene of the *E. coli* chromosome (O'Donnell Declaration ¶ 17). At the end of the sequence of the *rlpB* gene shown in Figure 7, the last 230 base pairs constitute a sequence that encodes the first 20-25% of the holA gene sequence (O'Donnell Declaration ¶ 17). Takase did not recognize this to be an open reading frame of a putative unknown gene, nor did the reference disclose the complete sequence of the gene (O'Donnell Declaration ¶ 17). The diagram attached as Exhibit 4 to the O'Donnell Declaration (diagram attached hereto as Exhibit 9) shows the overlap

between the disclosed *rlpB* gene of Takase and the holA gene encoding the claimed δ subunit (O'Donnell Declaration ¶ 17). Further, as shown in Dong, et al., "DNA Polymerase III Accessory Proteins," J. Biological Chem., 268(16):11758-765, 11759 n. 3 ("Dong"), Takase's published sequence was incorrect and incomplete, in fact, the first 54 nucleotides of the δ gene are incorrect by 11 nucleotides (O'Donnell Declaration ¶ 17). Thus, the δ protein subunit of polymerase III holoenzyme and the gene encoding the δ protein subunit of the polymerase III holoenzyme of the present invention are not disclosed by Takase.

Claim 54 relates to "[a]n isolated protein subunit of polymerase III holoenzyme, wherein the subunit group is δ." Further, claim 59 relates to "[a]n isolated DNA molecule encoding a protein subunit of polymerase III holoenzyme, wherein the subunit group is δ." Takase does not teach the specified isolated protein subunit of polymerase III holoenzyme, nor the gene encoding that protein. Further, Takase does not disclose the claimed expression system or host cell. Therefore, the rejection based on Takase is improper and should be withdrawn.

The rejection of claim 1 under 35 USC § 102(b) as anticipated by Stirling et al., "*xerB*, an *Escherichia coli* Gene Required For Plasmid ColE1 Site-Specific Recombination, Is Identical to *pepA*, Encoding Aminopeptidase A, a Protein With Substantial Similarity to Bovine Lens Leucine Aminopeptidase," The EMBO Journal 8:1623-27 (1989) ("Stirling") is respectfully traversed.

Stirling relates to the *xerB* gene from *E. coli* and the purification of the 55.3 kd *xerB* polypeptide. Figure 2(A) discloses the DNA sequence of the *xerB* gene which is located in a nucleotide sequence of a slightly larger fragment. Stirling discloses that the *xerB* gene is a 503-codon open reading frame which would encode a polypeptide of 55.3 kd. The position of the PTO is that this sequence matches that of the sequence encoding the claimed χ subunit. Stirling, however, does not disclose the polymerase III holoenzyme nor a

protein subunit of the polymerase III holoenzyme. In addition, Stirling only recognized the start of the gene open reading frame, but did not recognize the complete isolated DNA molecule or the complete protein it encodes.

Claim 43 relates to "[a]n isolated protein subunit of polymerase III holoenzyme, wherein the subunit group is χ." Further, claim 47 relates to "[a]n isolated DNA molecule encoding a protein subunit of polymerase III holoenzyme, wherein the subunit group is χ." Stirling does not teach the specified isolated protein subunit of polymerase III holoenzyme, nor the gene encoding that protein. Further, Stirling does not disclose the claimed expression system or host cell. Therefore, the rejection based on Stirling is improper and should be withdrawn

The rejection of claim 1 under 35 USC § 102(b) as anticipated by Yoshikawa et al., "Cloning and Nucleotide Sequencing of The Genes *rimI* and *rimJ* Which Encode Enzymes Acetylating Ribosomal Proteins S18 and S5 of *Escherichia coli* K12," Mol. Gen. Genet. 209:481-88 (1987) ("Yoshikawa") is respectfully traversed.

Yoshikawa relates to the *rimI* gene of *E. coli* which encodes for the RimI enzyme, which was deduced to contain 161 amino acid residues with a calculated molecular weight of 18,232. The nucleotide sequence of *rimI* is shown in Figure 3 of the reference. Yoshikawa did not recognize the segregated Ψ subunit or the segregated gene encoding the Ψ subunit. As pointed out in the present application, Yoshikawa ". . . does not indicate any appreciation of the gene as a coding sequence for the Ψ peptide" (page 56, lines 3, 4). The holD gene actually overlaps the disclosed *rimI* gene for 32 nucleotides. In other words, the last 32 nucleotides of holD (29 that encode amino acids, and 3 that encode the stop signal) are embedded within the first 32 nucleotides of the *rimI* protein disclosed in Yoshikawa. Yoshikawa does not, however, disclose this segregated sequence. In fact, Yoshikawa examined protein expression from the *rimI* gene, and did detect *rimI* protein, but failed to detect the segregated

holD (Ψ) protein expression even though the *rimI* gene contained the holD gene.

By contrast, claim 32 relates to "[a]n isolated segregated protein subunit of polymerase III holoenzyme wherein the subunit group is Ψ." Claim 36 relates to "[a]n isolated DNA molecule encoding a segregated protein subunit of polymerase III holoenzyme wherein the subunit group is Ψ ". None of these limitations are disclosed in Yoshikawa. Further, there is no disclosure of an isolated protein, "wherein the protein corresponds to an amino acid sequence corresponding to SEQ. ID. NO. 38" as recited in claim 33 or an isolated DNA molecule "wherein the DNA molecule corresponds to a nucleotide sequence corresponding to SEQ. ID. NO. 39" as recited in claim 37. There is also no disclosure of the claimed expression system and host cell. Thus, the rejection based on Yoshikawa is improper and should be withdrawn.

The provisional rejection of claim 1 under the judicially created doctrine of obvious-type double patenting is respectfully traversed in view of the cancellation of claim 1 in the preliminary amendment dated August 5, 1997. Accordingly, this rejection should be withdrawn.

In view of the foregoing, it is submitted that this case is in condition for allowance, and such allowance is earnestly solicited.

Respectfully submitted,

Date: January 5, 1998

Michael L. Goldman
Registration No. 30,727
Attorney for Applicant

Nixon, Hargrave, Devans & Doyle LLP
Clinton Square, P. O. Box 1051
Rochester, New York 14603
Telephone: (716) 263-1304
Telecopy: (716) 263-1600

© 1993 Oxford University Press

Nucleic Acids Research, 1993, Vol. 21, No. 1 1–3

Homology in accessory proteins of replicative polymerases—*E.coli* to humans

Mike O'Donnell[1,2], Rene Onrust[1], Frank B.Dean[3], Mei Chen[3] and Jerard Hurwitz[3]
[1]Microbiology Department and [2]Howard Hughes Medical Institute, Cornell University Medical College, 1300 York Avenue, New York, NY 10021 and [3]Program in Molecular Biology, Sloan-Kettering Institute, Memorial Sloan-Kettering Cancer Research Center, 1275 York Avenue/Box 97, New York, NY 10021, USA

Received November 4, 1992; Accepted November 19, 1992 GenBank accession nos*

The basis for the remarkably high processivity of DNA polymerases that duplicate long chromosomes appears quite similar in prokaryotes and eukaryotes. In each of these cell types, the replicative polymerase has several accessory proteins which endow the polymerase subunit with its speed and processivity. The replicative polymerases of the well studied systems of bacteriophage T4, *E.coli* (DNA polymerase III holoenzyme) and humans (polymerase δ), contain accessory proteins which form a 'sliding clamp' on DNA that acts to tether the polymerase to the DNA for rapid and highly processive synthesis (1–4). In the *E.coli* system this sliding clamp has been shown to be a dimer of the β subunit, which is in the shape of a ring encircling the DNA (5). The functional homologue of β in the T4 system is the product of gene 45 (g45 protein) and in humans it is the proliferating cell nuclear antigen (PCNA). These proteins are homologous in function, and although they show no homology at the amino acid sequence level, a case has recently been made for a structural similarity of β to PCNA and to the T4 g45 protein on the basis of sequence using the crystal structure of β as a guide (5).

These replicative polymerases each require several other accessory proteins to assemble the sliding clamp around the DNA. A list of these proteins is presented in Table I. The requirement for more than one protein in this reaction may reflect a need for several functions, including recognition of a primed template, the opening and closing of the ring shaped clamp protein around the DNA, and the coupling of this process to ATP hydrolysis. In the *E.coli* system this 'clamp loader' is the 5-protein γ complex ($\gamma\delta\delta'\chi\psi$), in humans it is the 5-protein Activator 1 (A1), also referred to as RF-C, and in phage T4 it is the 5 subunit g44 protein/g62 protein complex (the g44/62 complex).

The mechanism of the 'clamp loader' in these three systems may be similar. The sites of binding to the primer-template junction are similar for the *E.coli* γ complex and β subunits, the human A1 and PCNA, and the g44/62 and g45 proteins (8–11). In the case of *E.coli*, the loading of the β subunit forms a 'preinitiation complex', dependent upon ATP hydrolysis, that can be isolated by gel filtration (12,13). The action of the γ complex in this reaction can be substituted by pairs of the subunits, γδ, or τδ' (14). The preinitiation complex assembled by the yeast or human A1 and PCNA and dependent upon ATP hydrolysis can also be isolated by gel filtration (7,15). The complex assembled by the T4 proteins on primed DNA differs, in that while it can be cross-linked to the DNA, it is not stable enough to be isolated after gel filtration; a complex containing the g45 can only be isolated by gel filtration if the DNA polymerase (g43) is present with the g44/62 and ATP (9,16).

The ATPase activity of the multisubunit complex in each of the three systems is required to assemble the 'sliding clamp' and the polymerase onto the template, but is not required for the subsequent replication of long stretches of the template in vitro. This indicates that ATP hydrolysis by the accessory protein complex is not required for translocation by the polymerase, at least along a template unencumbered by other bound proteins.

Recently, the genes encoding most of these subunits have been identified. Comparison of the amino acid sequences in Fig. 1 shows several of these subunits are truly homologous between *E.coli*, humans and phage T4, especially in one region toward the middle (110–160), implying for the first time that the functional similarity among these systems has its basis in an evolutionarily conserved structure. Presumably all these proteins evolved from the same ancestral gene.

The homology between these protein subunits actually starts near the amino terminus, continues through the ATP-binding domain which starts approximately at the site of the 'GKT' box, and extends for about two hundred amino acid residues toward their carboxyl termini. All these proteins have a strictly conserved 'SRC' sequence of unknown function at residues 157–159. The presence of the four amino-acid 'DEAD' motif in three of the human protein subunits represent an unusual occurrence of this

Table I.

	E.coli	Eukaryotic	phage T4
DNA polymerase	Pol III core ($\alpha\epsilon\theta$)†	pol δ (pol ε)*	g43
Accessory complex	γ complex ($\gamma\delta\delta'\psi\chi$)	Activator 1 (A1) (aslo called RF-C)	g44/62
Accessory protein clamp	β	PCNA	g45

†The 10th subunit of *E.coli* DNA polymerase holoenzyme, τ, binds together two molecules of pol III core (24, 25) presumably for coordinated synthesis of both strands of duplex DNA.
*Polε is a different molecule from polδ and it is also activated by the eukaryotic accessory proteins (6,7).

*L04577, L07540, L07541

sequence in proteins not having RNA-dependent ATPase or putative RNA helicase activities. However, as mentioned above, ATP hydrolysis by these protein complexes is not required for translocation along the DNA template by the polymerase.

An interesting feature to emerge from the sequence comparisons is the similarity among subunits present within one polymerase accessory factor. In *E.coli*, the δ' subunit shares 27% identity to the γ and τ subunits; γ and τ are derived from the same gene and therefore have identical sequences (a translational frameshift produces γ (47 kDa) which is 24 kDa shorter than τ, ref. 23). The gene encoding δ' (*holB*) also produces two subunits, δ'_{small} and δ'_{large}, which differ by 520 da. δ'_{small} is the product of the full gene and therefore δ'_{large} is produced by a modification, possibly a translational frameshift. In humans, four subunits of the A1 are highly homologous to one another. Hence, the 'clamp loader' complexes of *E.coli* and humans appear to contain a family of subunits of similar structure within them. Although the phage T4 g44 protein is not homologous to g62 protein, the subunit composition of the g44/g62 complex is 4:1 (g44-to-g62) (17). Thus, the g44/62 complex also has structural redundancy and may be compared to the case of the human proteins, in which four different proteins having homology, the 40, 38, 37, and 36 kDa subunits, are present in the complex with one additional subunit, of 145 kDa, that has little homology to the other four.

Why there is such structural redundancy within these complexes is unclear at present. However, these sequence alignments underscore the basic similarity in structure and function of the cellular replicase accessory proteins across the evolutionary spectrum.

REFERENCES

1. Stukenberg, P.T., Studwell, P.S. and O'Donnell, M. (1991) *J. Biol. Chem.* 266, 11328–11334.
2. Hurwitz, J., Dean, F.B., Kwong, A.D. and Lee, S.-H (1990) *J. Biol. Chem.* 265, 18043–18046.
3. Nossal, N.G., and Alberts, B.M. (1983) In Bacteriophage T4 (Mathews, C.K., Kutter, E.M., Mosig, G., and Berget, P.B., eds) pp. 71–81, American Society for Microbiology, Washington, D.C.
4. Kornberg, A. and Baker, T. (1991) DNA Replication. New York, W.H. Freeman.
5. Kong, X.-P., Onrust, R. O'Donnell, M. and Kuriyan, J. (1992) *Cell* 69, 425–437.
6. Lee, S.-H, Pan, Z.-Q., Kwong, A.D., Burgers, P.M.J. and Hurwitz, J. (1991) *J. Biol. Chem.* 266, 22707–22717.
7. Burgers, P.M.J. (1991) *J. Biol. Chem.* 266, 22698–22706.
8. Munn, M.M. and Alberts, B.M. (1991) *J. Biol. Chem.* 266, 20024–20033.
9. Capson, T.L., Benkovic, S.J. and Nossal, N.G. (1991) *Cell* 65, 249–258.
10. Tsurimoto, T. and Stillman, B. (1991) *J. Biol. Chem.* 266, 1950–1960.
11. Griep, M.A. and McHenry, C.S. (1992) *J. Biol. Chem.* 267, 3052–3059.
12. Wickner, S.H. (1976) *Proc. Natl. Acad. Sci. USA* 73, 3511–3515.
13. O'Donnell, M. (1987) *J. Biol. Chem.* 262, 16558–16565.
14. O'Donnell, M. and Studwell, P.S. (1990) *J. Biol. Chem.* 265, 1179–1187.
15. Lee, S.-H. and Hurwitz, J. (1990) *Proc. Natl. Acad. Sci. USA* 87, 5672–5676.
16. Richardson, R.W., Ellis, R.L. and Nossal. N.G. (1990) *UCLA Symp. Mol. Cell. Biol.* 127, 247–259.
17. Jarvis, T.C., Paul, L.S. and von Hippel, P.H. (1989) *J. Biol. Chem.* 264, 12709–12716.
18. Flower, A.M. and McHenry, C.S. (1986) *Nuc. Acids Res.* 14, 6541–6549.
19. Blinkowa, A. and Walker, J. (1986) *Nuc. Acids Res.* 14, 6541–6549.
20. Chen, M, Pan, Z.-Q and Hurwitz, J. (1992) *Proc. Natl. Acad. Sci USA* 89, 2516–2520.
21. Chen, M., Pan, Z.-Q and Hurwitz, J. (1992) *Proc. Natl. Acad. Sci. USA* 89, 5211–5215.
22. Spicer, E.K., Nossal, N.G. and Williams, K.R. (1984) *J. Biol. Chem.*, 259, 15425–15432.
23. Tsuchihashi, Z. and Kornberg, A. (1990) *Proc. Natl. Acad. Sci. USA* 87, 2516–2520.
24. McHenry, C.S. (1982) *J. Biol. Chem.* 257, 2657–2663.
25. Studwell-Vaughan, P.S. and O'Donnell, M (1991) *J. Biol. Chem.* 266, 19833–19841.

The EMBO Journal vol.14 no.16 pp.3966–3976, 1995

Rules governing the efficiency and polarity of loading a tracking clamp protein onto DNA: determinants of enhancement in bacteriophage T4 late transcription

Glenn M.Sanders, George A.Kassavetis and E.Peter Geiduschek

Department of Biology and Center for Molecular Genetics, University of California, San Diego, La Jolla, CA 92093-0634, USA

The bacteriophage T4 DNA polymerase accessory proteins confer processivity and high speed on replicative DNA chain elongation: the gene 45 protein, gp45, tracks along DNA and serves as the sliding clamp of the viral DNA polymerase; the gene 44/62 protein complex, gp44/62, is an ATP-dependent loading enzyme that mounts gp45 on DNA. Gp45 also activates T4 late transcription. Transcriptional enhancement by gp45 requires a particular orientation that is imposed by gp44/62 at the DNA loading site. Loading and orienting gp45 on DNA, tracking along DNA and interaction with RNA polymerase have been analyzed by measuring transcriptional activation. The efficiency of loading gp45 at different DNA structures and the resulting transcriptional activation have been compared, and sources of interference with transcriptional activation have been examined. All observations are compatible with a mechanism in which the loading enzyme recognizes the polarity of single-stranded DNA and imposes a corresponding polarity of DNA entry on gp45. Primer–template junctions are the most efficient DNA loading sites for gp45 and can generate very rapid opening at promoters that are located at a distance of >1 kbp. In contrast, gp45 does not track efficiently across single-stranded DNA.

Keywords: DNA tracking proteins/enhancers/gene regulation/processivity factor/T4 late genes

Introduction

The primary replicative DNA polymerases of eukaryotes, prokaryotes and certain viruses share common elements of structure and function: they are composed of a core enzyme, capable of relatively non-processive DNA synthesis and exonucleolytic degradation, and a set of accessory proteins that confer processivity on the core enzyme (Kornberg and Baker, 1992; Kuriyan and O'Donnell, 1993; Nossal, 1994). The accessory proteins comprise two components: a processivity factor proper, the 'sliding clamp', and an ATP-dependent assembly factor, the 'clamp loader'. The structures of two sliding clamps, the β subunit of *Escherichia coli* DNA polymerase III holoenzyme and yeast proliferating cell nuclear antigen (PCNA), have been determined: they are toroidal multimers, capable of encircling DNA (Kong *et al.*, 1992; Krishna *et al.*, 1994). Indeed, eukaryotic (PCNA), bacterial (β) and viral [bacteriophage T4 gene 45 protein (gp45)] clamps have been shown to 'track', that is to diffuse freely on DNA although confined topologically, once they have been loaded onto DNA at an appropriate site by the cognate 'clamp loader' (Stukenberg *et al.*, 1991; Burgers and Yoder, 1993; Tinker *et al.*, 1994b; Podust *et al.*, 1995). How the clamp loaders, clamps and DNA loading sites interact can also be understood through analysis of DNA replication and through measurements of the ATPase activities of these clamp loaders (reviewed by Young *et al.*, 1992, 1994; Nossal, 1994).

A role for one set of DNA polymerase accessory proteins in transcriptional activation has been identified recently, permitting examination of accessory protein loading and tracking in the context of transcription. The late genes of bacteriophage T4 are transcribed from extremely simple promoters, consisting of TATAAATA centered ~10 bp upstream of a transcriptional start site. The ability to recognize these promoters is conferred on *E.coli* RNA polymerase core by a small σ-family protein encoded by T4 gene 55 (gp55), but transcription is weak, particularly in relaxed or linear DNA, and further suppressed by the small RNA polymerase-bound T4 gene 33 protein. Gp33 is a transcriptional co-activator; it confers the ability to support activation of transcription by gp45, the 'sliding clamp' of the T4 DNA polymerase holoenzyme (Herendeen *et al.*, 1990). Although gp45 alone can activate gp55-directed transcription in the presence of gp33 under conditions of macromolecular crowding (Sanders *et al.*, 1994), under conventional reaction conditions it additionally requires the 'clamp loader' encoded by T4 genes 44 and 62 (the gp44/62 complex). The latter loads gp45 onto DNA in an ATP hydrolysis-dependent process at enhancer-like entry sites that can be located at a considerable distance from the promoter (reviewed by Brody *et al.*, 1995).

Once gp45 has been loaded, it tracks along DNA, encounters RNA polymerase, and eventually becomes stably associated with the upstream end of the activated open complex (Tinker *et al.*, 1994a). Tracking along DNA is an essential part of this transcriptional activation mechanism: a continuous and open path along DNA connecting an enhancer and its promoter is required for activation by the T4 DNA polymerase accessory proteins (Herendeen *et al.*, 1992).

The most studied enhancers of T4 late transcription have been nicks in DNA, which can activate transcription from upstream or downstream of a target promoter. A characteristic polarity distinguishes these *cis*-acting sites from conventional enhancers: the nick must be in the non-transcribed strand of its target transcription unit (Herendeen *et al.*, 1989). It has been suggested that polarity of transcriptional activation is due to assembly of an asymmetric gp44/62–gp45 complex at the nick-as-enhancer, imposing a particular orientation on gp45 as it tracks along DNA and thereby determining the orientation of RNA polymerase that is compatible with formation of

© Oxford University Press

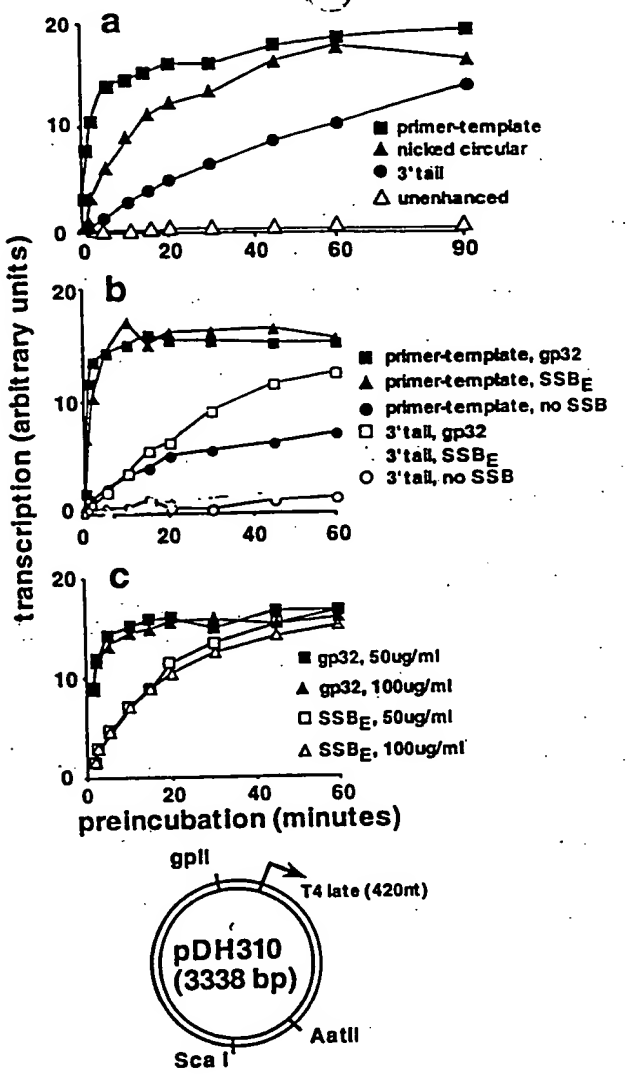


Fig. 2. Efficiencies of transcriptional activation generated by different gp45 loading sites. A mixture of T4-modified RNA polymerase, gp55, gp33 and the DNA polymerase accessory proteins (gp44/62 complex and gp45) was added to pDH310 (shown at the bottom of the figure) DNA, gp32 and dATP pre-equilibrated at 25°C. Aliquots were withdrawn at the times noted on the abscissa and added to a mixture of unlabeled and radioactive ribonucleoside triphosphates and rifampicin pre-equilibrated at 25°C for a single round of transcription. The yield of 420 nt RNA is reported on the ordinate (in arbitrary units). Further experimental details are provided in Materials and methods. (a) Comparison of primer–template and 3′-tailed junctions with nicked circular DNA. (■): linear DNA with 5′ overhanging (primer–template junction) ends; (●): linear DNA with 3′-tailed ends; (▲): nicked circular DNA; (△): unenhanced transcription (accessory proteins omitted). The distances from the relevant gp45 loading sites (see below) to the enhanced promoter are ~1520 bp, ~1070 bp and ~220 bp for the primer–template junction, 3′ tail and nick, respectively. (b) Effects of gp32 and *E.coli* SSB (SSB$_E$) on the efficiency of transcriptional activation. DNA with 5′ overhanging (closed symbols) and 3′ overhanging (open symbols) ends, without SSB (circles), with SSB$_E$, (triangles) or with gp32 (squares) was analyzed. (c) Effects of gp32 and SSB$_E$ on activation of transcription of DNA with 5′ overhanging ends, at one-fourth of the concentration of each of the DNA polymerase accessory proteins used in (b). Gp32 (closed symbols) and SSB$_E$ (open symbols) were used at 100 μg/ml (triangles) or 50 μg/ml (squares).

the transcription initiation inhibitor rifampicin. The assay measures the end-product of a complex reaction sequence, which requires an adequate measure of thermal equilibration, assembly of a gp44/62–gp45 complex at a loading site, entry of gp45 for tracking along DNA, interaction with RNA polymerase, promoter location and opening. Nevertheless, primer–template junctions generate half-maximal promoter opening within 1–2 min, at modest concentrations of the DNA polymerase accessory proteins (Figure 2a, closed squares). Promoter opening of nicked circular templates and of templates bearing recessed 5′ ends (i.e. 3′ tailed) is relatively slow, with 50% rise times of ~10 min and >30 min, respectively (Figure 2a, closed triangles and closed circles, respectively). Unenhanced transcription is negligible under these conditions (open triangles). Thus, the most effective of these enhancers (see below) is a primer–template junction that is located ~1.5 kbp from the activated promoter.

SSB proteins affect the ability of single-stranded DNA regions to generate enhanced transcription

The preceding experiments were carried out in the presence of gp32, the T4-encoded SSB. The effect of omitting gp32 or of substituting the *E.coli* SSB (SSB$_E$) on activation of transcription has also been examined (Figure 2b). In the absence of any SSB, 3′-tailed DNA does not generate activated transcription (open circles), and SSB$_E$ does not rescue activation (open triangles), implying that the lack of transcription is not merely due to sequestration of transcription components by single-stranded DNA. Significant rates of promoter opening with 3′ tails are only obtained in the presence of gp32 (open squares). In contrast, primer–template junctions generate modest rates of promoter opening in the absence of any SSB (closed circles), and the rate of promoter opening is greatly increased by providing SSB$_E$ (closed triangles). Gp32 further increases the rate of promoter opening (closed squares). SSB$_E$ and gp32 have no effect on transcription of nicked, circular DNA (data not shown). Thus, a single-stranded region of DNA is necessary for the SSBs to manifest their effects on activation of transcription.

The profound effect of both SSB$_E$ and gp32 on enhancement of transcription from a primer–template junction implies that simply masking exposed single-stranded DNA greatly improves loading of gp45. The next experiment (Figure 2c) looks for a more specific role of SSB in gp45 loading by examining quantitative differences in the ability of gp32 and SSB$_E$ to facilitate transcriptional enhancement. Since SSB$_E$ and gp32 differ greatly in their modes of DNA association (Lohman and Ferrari, 1994), differences in transcriptional activation could also reflect different degrees of saturation of single-stranded DNA. Rates of promoter opening were examined at limiting concentrations of the DNA polymerase accessory proteins (in order to slow down the otherwise rapid rate of open complex formation and accentuate differences between gp32 and SSB$_E$ in promoting enhanced transcription). At their respective saturation limits (excluding the possibility that quantitative differences in activation reflect different degrees of saturation of single-stranded DNA with SSB), gp32 is a substantially more effective co-factor than SSB$_E$ for enhancement of transcription from a primer–template junction (Figure 2c, compare closed symbols with open

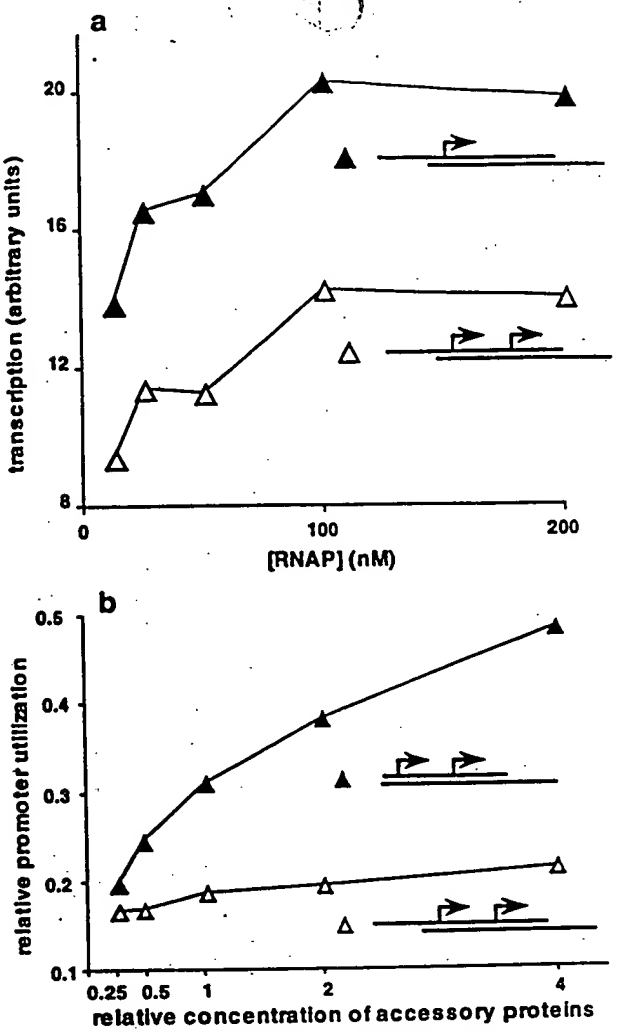


Fig. 4. Downstream promoters and upstream loading sites can interfere with activation of transcription. (a) Single-round transcription of linear pDH72ΔE1 DNA (Figure 3a) with 5′ overhanging ends, and two T4 late transcription units in tandem (open triangles), compared with a similar template (pDH72Δ420) from which the right (420 nt) promoter has been deleted (filled triangles). The yields of 314 nt RNA from the left promoters are compared as a function of RNA polymerase concentration. (b) Single-round transcription of linear pDH72ΔE1 with primer–template junction loading sites for gp45 at both ends (open triangles), or only at the activating downstream end (filled triangles). The ratio of the molar yield of left promoter-derived RNA (314 nt) to right promoter-derived RNA (420 nt) is compared as a function of DNA polymerase accessory protein concentration, one unit corresponding to 160 nM gp 44/62 complex with 90 nM gp45.

from its downstream, but not its upstream, gp45 loading site. Plasmid pGS724 has two divergent T4 late transcription units, separated by a single *Eco*RI site. Addition of *Eco*RI-Gln111 to this template isolates each promoter from its upstream, but not its downstream, enhancer (Figure 3b and c, top).

Primer–template junctions and 3′ tails at both ends of pGS722 are compatible with activation of transcription at both promoters (Figure 3b and c, lanes 2). Addition of *Eco*RI-Gln111 abolishes enhanced transcription from both promoters, regardless of whether the loading site is a primer–template junction (Figure 3b, lanes 3–5) or a 3′ tail (Figure 3c, lanes 3–5). Activation of transcription on pGS724 is also effective at both promoters (Figure 3b and c, lanes 7), but addition of *Eco*RI-Gln111 to pGS724 has little effect on enhanced transcription, again regardless of whether the loading site is a primer–template junction (Figure 3b, lanes 8–10) or a 3′ tail (Figure 3c, lanes 8–10). That a single *Eco*RI site is sufficient to block DNA tracking by gp45 has been shown previously (Herendeen *et al.*, 1992) and is verified here by the disappearance of an enhancement-dependent, higher molecular weight transcript upon addition of *Eco*RI-Gln111 (data not shown). We conclude that activation of transcription from a single-stranded DNA end only occurs from downstream of a T4 late promoter, and can be generated from primer–template junctions or 3′ tails.

In contrast, the results of Figure 1 suggest that only promoters with single-stranded extensions on their transcribed strands are activated (lanes 11–14, in particular). A solution to this apparent conflict lies in the orientation of the promoters of pDH82, the requirement for a clear and unobstructed pathway between loading site and promoter, and the relative efficiency of a primer template junction as an activator of transcription. Activation of each of the convergent promoters of pDH82 (Figure 1) by a downstream loading site necessitates gp45 tracking past the other promoter. Open promoter complexes should form roadblocks to enhancement of other promoters in *cis*. In the case of pDH82 (Figure 1) bearing an efficient primer–template junction loading site for gp45 at one end, and a relatively weak 3′ tail loading site at the other ('top-long' and 'bottom-long' DNA), the promoter upstream of the efficient loading site is expected to open quickly (Figure 2) and block enhancement of the promoter upstream of the inefficient loading site.

Traffic problems on the track: upstream enhancers and downstream promoters can interfere with activation

The preceding supposition has been tested directly by comparing open promoter complex formation on pDH72ΔE1 (Figure 4, top) with open promoter complex formation on a derivative template, pDH72Δ420, from which the potentially interfering downstream promoter has been deleted. Indeed, the yield of transcripts is increased by deletion of the downstream promoter (Figure 4a), at limiting or excess RNA polymerase, and is therefore unlikely simply to reflect competition for RNA polymerase by the additional late promoter on pDH72ΔE1.

To explain more fully the weak activation of the upstream promoter in Figure 3, we examined whether gp45 loaded at a non-activating site (upstream of a proximal promoter, the stream referring to the direction of transcription) can interfere with activation from a gp45 loading site located downstream of that promoter. The effect of accessory protein concentration on the relative yield of transcripts from tandem promoters was compared on DNA with a gp45 loading site only at the downstream end (the upstream end having been made blunt), and on DNA with gp45 loading sites at both ends, placing a non-activating gp45 loading site upstream of the left-hand promoter (Figure 4b). Increasing the accessory protein concentration increases the relative yield of transcripts from the left promoter on the template with only a downstream gp45 loading site (filled triangles), but not on the template with gp45 loading sites at both ends (open triangles). These results are consistent with the supposition

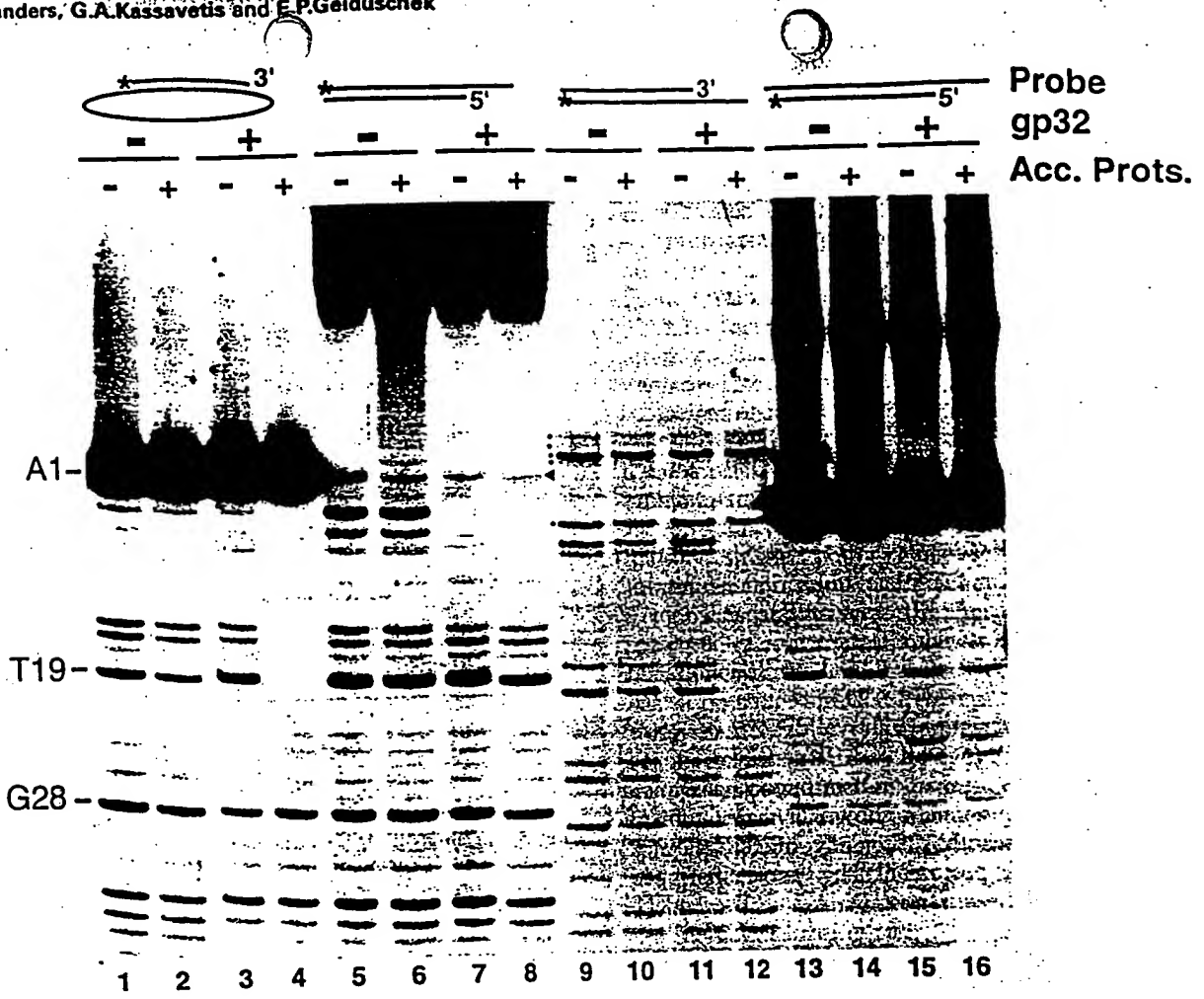


Fig. 6. DNase I footprints of protein complexes at double strand–single strand junctions. Probes for DNase I footprinting, diagrammed at the top of the figure, were incubated with the indicated proteins, in the presence of 200 μM ATP-γ-S, then digested with DNase I, and analyzed as described in Materials and methods. Small amounts of labeled fragments contaminating probes for lanes 5–8 and 9–12 are marked with a triangle and black dots, respectively, and are excluded from the analysis. Lanes 1–4: primer–template junction probe, 5′ end-labeled in the short strand; lanes 5–8: 3′-tailed DNA labeled in the long strand; lanes 9–12: primer–template junction probe labeled in the short strand; lanes 13–16: 3′-tailed DNA labeled in the short strand. Lanes 1, 5, 9 and 13: no-protein controls; lanes 2, 6, 10 and 14: with DNA polymerase accessory proteins alone; lanes 3, 7, 11 and 15: with gp32 alone; lanes 4, 8, 12 and 16: with accessory proteins and gp32.

that allows partial invasion of the duplex region. The failure to protect the complementary short strand is consistent with the high cooperativity of gp32 binding to single-stranded DNA. The length of single-stranded DNA that is exposed at the end of the complementary DNA strand would accommodate only one molecule of gp32, which would bind relatively inefficiently (Kowalczykowski *et al.*, 1981). Nevertheless, melting of this short section of the 5′-recessed strand at the double strand–single strand junction could render it capable of being loaded with gp45 in an orientation that is compatible with transcriptional activation. The efficiency of gp45 loading at this site is at least an order of magnitude lower than at the primer–template junction (Figure 2a), consistent with the lack of a clear-cut footprint of DNA polymerase accessory proteins in lanes 8 and 16.

Discussion

A model to explain the transcriptional enhancement properties of diverse DNA structures

Nicks in DNA, gaps and single-stranded extensions of either polarity all serve as loading sites for the DNA polymerase processivity factor and transcriptional activator gp45, but with greatly different efficiencies (Figures 1 and 2). Interference with transcriptional activation can arise in two ways: gp45 tracking along DNA in the activation-incompatible orientation diminishes the activity of properly oriented gp45 (Figure 4b), and open promoter complexes can block the track to gp45 (Figure 4a).

The seemingly disparate properties of different gp45 loading sites and of differently organized and oriented transcription units can be reconciled by a single model with the following features (Figure 7): (i) the loading enzyme, gp44/62, (Kaboord and Benkovic, 1995) recognizes the 5′→3′ polarity of the continuous DNA strand at the loading site. (ii) The gp44/62–single-stranded DNA interaction determines the polarity of loading of gp45. (iii) Once the orientation of gp45 on DNA is determined, it is not reversed subsequently (Tinker *et al.*, 1994a). (iv) When gp45 enters its complex with gp44/62 at the DNA loading site, it is situated on the side of gp44/62 that faces the 3′ end of the continuous or longer DNA strand. At a primer–template junction, this places gp45 over double-stranded DNA (Figure 7, line b), consistent with the prior footprinting and photocrosslinking analyses

orientation of gp45 loading is established by the 5′→3′ polarity of the continuous strand, which must be the transcribed strand of the activated promoter. We conclude that gp45 is first released from its loading site at the DNA nick toward the left, as drawn on Figure 7, line g, because converting the nick into an ~250 nt gap confines activation by gp45 to that side (line h and Figure 5).

The known properties of the T4 DNA polymerase accessory proteins are consistent with this model. Single-stranded DNA and primer–template junctions are equally effective co-factors for the DNA-dependent ATPase activity of gp44/62 complex alone; a preference for the primer–template junction is conferred by gp45 (Jarvis *et al.*, 1989b). A photocrosslinking analysis of the DNA polymerase accessory protein complex at a primer–template junction places gp45 on double-stranded DNA (Capson *et al.*, 1991), consistent with the supposition that the specific preference of the gp44/62–gp45 complex for a primer–template junction is conferred by DNA duplex-confined gp45 stabilizing the placement of gp44/62 complex on single-stranded DNA. The cellular counterparts of gp45, the β subunit of *E.coli* DNA polymerase III holoenzyme and PCNA are tori (Kong *et al.*, 1992; Krishna *et al.*, 1994). Anticipating that gp45 will also turn out to be toroidal, one can rationalize readily that the central hole of gp45 might be too small for tracking along gp32-laden single-stranded DNA and that DNA secondary structure, as well as direct interactions of nucleotides in single-stranded DNA with the external faces of the protein catenane, would also obstruct tracking along bare single-stranded DNA. We do not know the size of the smallest gap that forms an effective barrier to gp45 tracking, but would not be surprised to find gp45 able to cross a short gap such as might be created at an Okazaki fragment junction by digestion of the RNA primer (Nossal, 1994). If the gp45 trimer (Jarvis *et al.*, 1989a) is a PCNA-like torus with a 3-fold axis down its central cavity (Krishna *et al.*, 1994), then its lateral faces must be non-identical, capable of different specific protein–protein interactions and of manifesting the polarity that is required to account for the properties of transcriptional enhancement at late promoters exhibited in this and prior work (Herendeen *et al.*, 1989; Tinker *et al.*, 1994a).

Efficiencies of transcriptional activation

Extraordinarily effective transcriptional activation, reflected in very rapid promoter opening, is afforded by primer–template junctions in the presence of gp32 (Figure 2). Activation, as measured by the rate of accumulation of open promoter complexes, is relatively inefficient in the absence of any SSB (Figure 2), probably due to non-productive sequestration of proteins on single-stranded DNA tails. This can be prevented by adding (the heterologous) SSB_E. SSBs probably also prevent tracking gp45 from falling off the ends of linear DNA. The effect is comparable with that of confining tracking gp45 by using a circular DNA template, or by blocking the ends of its linear DNA with tightly bound protein (cf. Stukenberg *et al.*, 1991; Herendeen *et al.*, 1992).

There is, in addition, a specific quantitative effect of gp32 on the enhancer efficiency of primer–template junctions (Figure 2) and an absolute requirement for gp32 that is not filled by SSB_E when recessed 5′ DNA ends serve as gp45 loading sites. We think it plausible to attribute these effects to specific interactions of gp32 with gp45 as well as with gp44/62 complex (Formosa *et al.*, 1983). That gp32 strongly increases the affinity of gp44/62 complex and gp45 for a primer–template junction, as judged by footprinting and gel filtration (Munn and Alberts, 1991a; Richardson *et al.*, 1989), is probably due to these interactions. Photocrosslinking experiments have shown recently that specific interaction with gp32 increases the density of gp45 tracking on DNA, and that the general non-specific DNA binding activity of SSB_E or the human SSB, RPA, cannot substitute for gp32 in this regard (Tinker *et al.*, 1994b).

Since the rate of open promoter complex formation can be extremely rapid, and is strongly dependent on the properties of the gp45 loading site, events subsequent to gp45–RNA polymerase encounter must also be very rapid. We infer that differences in rates of open promoter complex accumulation primarily reflect differences in rates of gp45 loading (and possibly of unloading, which would also affect the density of tracking gp45). This view is consistent with the observed hierarchies of transcriptional activation. Optimal activation of transcription is achieved with primer–template junctions; nicks and 3′ tails are less efficient (Figure 2), because they are sub-optimal as binding sites for the gp44/62 complex (e.g. Figure 6) and consequently load gp45 less efficiently. Current experiments (T.-J.Fu, personal communication) are directed at exploring the dynamics of gp45 loading and tracking.

In closing, we want to point out that the most efficient enhancer in this resolved, highly purified and simplified *in vitro* system is not necessarily the primary contributor to T4 late transcription *in vivo*. Primer–template junctions are sites of assembly of DNA polymerase holoenzyme and of active DNA chain elongation, both expected to compete with gp45 loading. (Preliminary experiments confirm, for example, that stalled DNA polymerase holoenzyme blocks T4 late transcriptional enhancement by gp45; G.M.S. unpublished observations.) On the other hand, rapid promoter opening (Figure 2) allows a variety of gp45 loading sites to contribute incrementally to T4 late gene activity, and may also permit a single gp45 trimer to activate several rounds of transcription initiation during a single episode of tracking along DNA.

Materials and methods

Labeled and unlabeled nucleoside triphosphates, bovine serum albumin (BSA), terminal deoxynucleotidyl transferase (TdT), exonuclease III (exo III), DNA ligase, DNase I, proteinase K, restriction enzymes and rifampicin were purchased from various commercial suppliers.

Plasmids pDH82, pDH310 and the pDH72 series have been described (Herendeen *et al.*, 1989, 1992; Herendeen, 1991). Plasmid pGS722 is a derivative of pDH72ΔE1 in which the 420 nt T4 late transcription unit was inverted by excision with *Aat*II and *Nsi*I restriction enzymes, digestion of the 3′ overhanging ends with T4 DNA polymerase, and religation of the fragments. Plasmid pGS724 was generated by inserting the *Sma*I–*Acc*I fragment of pTE110 containing the 420 nt T4 late transcription unit, into the *Sma*I site of pTE114 (Elliot and Geiduschek, 1984). Plasmid pDH72Δ420 (Figure 4a) was generated by deleting the 150 bp *Nsi*I–*Dra*III fragment of pDH72, which contains the 'right' T4 late promoter that yields 420 nt RNA, but retaining the T4 late transcription unit that yields 314 nt RNA. Plasmid pGS725 was generated by adding *Eco*RI linkers to the *Aff*III and *Eco*109 sites of pDH72ΔE123. (Complete sequences of these plasmids are available from the authors on request.)

Capson,T.L., Benkovic,S.J. and Nossal,N.G. (1991) Protein–DNA cross-linking demonstrates stepwise ATP-dependent assembly of T4 DNA polymerase and its accessory proteins on the primer–template. *Cell*, 65, 249–258.

Elliott,T. and Geiduschek,E.P. (1984) Defining a bacteriophage T4 late promoter: absence of a '–35' region. *Cell*, 36, 211–219.

Formosa,T., Burke,R.I. and Alberts,B.M. (1983) Affinity purification of bacteriophage T4 proteins essential for DNA replication and genetic recombination. *Proc. Natl Acad. Sci. USA*, 80, 2442–2446.

Gogol,E.P., Young,M.C., Kubasek,W.L., Jarvis,T.C. and von Hippel,P.H. (1992) Cryoelectron microscopic visualization of functional sub-assemblies of the bacteriophage T4 DNA replication complex. *J. Mol. Biol.*, 224, 395–412.

Goodrich,L.D., Lin,T.-C., Spicer,E.K. and Konigsberg,W.H. (1992) Effects of deletion of the carboxy-terminus of T4 DNA polymerase upon its interaction with its accessory proteins. *J. Cell Biochem.*, 16B, 45.

Herendeen,D.R. (1991) Regulation of bacteriophage T4 late gene transcription. Ph.D. Thesis, U.C. San Diego.

Herendeen,D.R., Kassavetis,G.A., Barry,J., Alberts,B.M. and Geiduschek,E.P. (1989) Enhancement of bacteriophage T4 late transcription by components of the T4 DNA replication apparatus. *Science*, 245, 952–958.

Herendeen,D.R., Williams,K.P., Kassavetis,G.A. and Geiduschek,E.P. (1990) An RNA polymerase-binding protein that is required for communication between an enhancer and a promoter. *Science*, 248, 573–578.

Herendeen,D.R., Kassavetis,G.A. and Geiduschek,E.P. (1992) A transcriptional enhancer whose function imposes a requirement that proteins track along DNA. *Science*, 256, 1298–1303.

Jarvis,T.C., Paul,L.S. and von Hippel,P.H. (1989a) Structural and enzymatic studies of the T4 DNA replication system. I. Physical characterization of the polymerase accessory protein complex. *J. Biol. Chem.*, 264, 12709–12716.

Jarvis,T.C., Paul,L.S., Hockensmith,J.W. and von Hippel,P.H. (1989b) Structural and enzymatic studies of the T4 DNA replication system. II. ATPase properties of the polymerase accessory protein complex. *J. Biol. Chem.*, 264, 12717–12729.

Kaboord,B.F. and Benkovic,S.J. (1995) Accessory proteins function as matchmakers in the assembly of the T4 DNA polymerase holoenzyme. *Curr. Biol.*, 2, 149–157.

Kong,X.P., Onrust,R., O'Donnell,M. and Kuriyan,J. (1992) Three-dimensional structure of the beta subunit of *E.coli* DNA polymerase III holoenzyme: a sliding DNA clamp. *Cell*, 69, 425–437.

Kornberg,A. and Baker,T.A. (1992) *DNA Replication*, 2nd edn., W.H. Freeman, New York.

Kowalczykowski,S.C., Lonberg,N., Newport,J.W. and von Hippel,P.H. (1981) Interactions of bacteriophage T4-coded gene 32 protein with nucleic acids. I. Characterization of the binding interactions. *J. Mol. Biol.*, 145, 75–104.

Krishna,T.S.R., Kong,X.-P., Gary,S., Burgers,P.M. and Kuriyan,J. (1994) Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell*, 79, 1233–1243.

Kuriyan,J. and O'Donnell,M. (1993) Sliding clamps of DNA polymerases. *J. Mol. Biol.*, 234, 915–925.

Lohman,T.M. and Ferrari,M.E. (1994) *Escherichia coli* single-stranded DNA-binding protein: multiple DNA-binding modes and cooperativities. *Annu. Rev. Biochem.*, 63, 527–570.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Munn,M.M. and Alberts,B.M. (1991a) The T4 DNA polymerase accessory proteins form an ATP-dependent complex on a primer–template junction. *J. Biol. Chem.*, 266, 20024–20033.

Munn,M.M. and Alberts,B.M. (1991b) DNA footprinting studies of the complex formed by the T4 DNA polymerase holoenzyme at a primer–template junction. *J. Biol. Chem.*, 266, 20034–20044.

Nossal,N. (1994) The bacteriophage T4 DNA replication fork. In Karam,J.D. (ed.), *Molecular Biology of Bacteriophage T4*. ASM Press, Washington, DC, pp. 43–53.

Podust,L.M., Podust,V.N., Sogo,J.M. and Hübscher,U. (1995) Mammalian DNA polymerase accessory proteins: Analysis of replication factor C—catalyzed proliferating cell nuclear antigen loading onto circular double-stranded DNA. *Mol. Cell. Biol.*, 15, 3072–3081.

Richardson,R.W., Ellis,R.L. and Nossal,N.G. (1989) In Richardson,C. and Lehman,R. (eds), *Molecular Mechanisms in DNA Replication and Recombination. UCLA Symposia on Molecular and Cellular Biology*, New Series, Alan R. Liss, Inc., NY, Vol. 127, pp. 247–259.

Sanders,G.M., Kassavetis,G.A. and Geiduschek,E.P. (1994) Use of a macromolecular crowding agent to dissect interactions and define functions in transcriptional activation by a DNA-tracking protein: bacteriophage T4 gene 45 protein and late transcription. *Proc. Natl Acad. Sci. USA*, 91, 7703–7707.

Stahl,S., Hultman,T., Olsson,A., Moks,T. and Uhlen,M. (1988) Solid phase DNA sequencing using the biotin–avidin system. *Nucleic Acids Res.*, 16, 3025–3038.

Stukenberg,P.T., Studwell-Vaughan,P.S. and O'Donnell,M. (1991) Mechanism of the sliding beta-clamp of DNA polymerase III holoenzyme. *J. Biol. Chem.*, 266, 11328–11334.

Tinker,R.L., Williams,K.P., Kassavetis,G.A. and Geiduschek,E.P. (1994a) Transcriptional activation by a DNA-tracking protein: structural consequences of enhancement at the T4 late promoter. *Cell*, 77, 225–237.

Tinker,R.L., Kassavetis,G.A. and Geiduschek,E.P. (1994b) Detecting the ability of viral, bacterial and eukaryotic replication proteins to track along DNA. *EMBO J.*, 13, 5330–5337.

Young,M.C., Reddy,M.K. and von Hippel,P.H. (1992) Structure and function of the bacteriophage T4 DNA polymerase holoenzyme. *Biochemistry*, 31, 8675–8690.

Young,M.C., Reddy,M.K., Jarvis,T.C., Gogol,E.P., Dolejsi,M.K. and von Hippel,P.H. (1994) Protein–protein and protein–DNA interactions in the T4 DNA polymerase accessory protein complex. In Karam,J.D. (ed.), *Bacteriophage T4*. ASM Press, Washington, DC, pp. 313–317.

Received on May 18, 1995; revised on June 14, 1995

Cell, Vol. 78, 725–728, September 9, 1994. Copyright © 1994 by Cell Press

Smart Machines at the DNA Replication Fork

Minireview

Bruce Stillman
Cold Spring Harbor Laboratory
Cold Spring Harbor, New York 11724

The protein apparatus that functions to replicate DNA has been likened to an efficient machine that rapidly and accurately duplicates genetic information for the next generation. Recent studies on both Escherichia coli and bacteriophage T4 not only reinforce this idea, but demonstrate that the replication machinery can monitor progress, cope with different molecular situations, and detect problems.

The Prokaryote Fork

The proteins that replicate the E. coli and bacteriophage T4 genomes have been identified and characterized in detail (see Table 1; Kornberg and Baker, 1992). It is not surprising that for such a fundamental process as replication of DNA, the functions of the individual proteins have been conserved, even in eukaryotes (see below). It is a little surprising, however, that in many cases, but not all, there is no obvious similarity between the amino acid sequences in proteins of the same functional class (Kong et al., 1992; O'Donnell et al., 1993). For example, the E. coli single-stranded DNA-binding protein (SSB), the phage T4 gene 32-encoded protein (32 protein), and the human replication protein A (RPA) are functionally similar, yet sequence unrelated. The same is true for the β subunit of polymerase III (pol III) from E. coli, the 45 protein from T4, and the human proliferating cell nuclear antigen (PCNA).

At the prokaryote DNA replication fork, a DNA helicase (DnaB or 41 protein) precedes the DNA synthetic machinery and unwinds the duplex parental DNA in cooperation with the SSB (see Figure 1A). Because of the complementary and antiparallel nature of the DNA double helix, DNA replication of the two strands occurs in a fundamentally different manner (see Figure 1A). On one strand, the leading strand, replication occurs continuously in a 5′ to 3′ direction, whereas on the other strand, the lagging strand, DNA replication occurs discontinuously by synthesis and joining of short Okazaki fragments. The leading-strand replication apparatus consists of a DNA polymerase (pol III core or 43 protein), a "sliding clamp" (β or 45 protein), and "brace" proteins (γ.δδ′χψ [the γ complex] or 44/62 proteins). All these proteins cooperate to replicate DNA in a rapid and processive manner, as discussed below. For example, the polymerase holoenzymes can easily synthesize greater than 50,000 nucleotides at more than 500 nt per second from a single primer without dissociating from the DNA template.

The brace proteins load the sliding clamp and the DNA polymerase onto the primer, either the initial primer at the origin of DNA replication during initiation of leading-strand synthesis or each RNA primer for Okazaki fragment synthesis (Capson et al., 1991; Munn and Alberts, 1991; Stukenberg et al., 1991; Gogol et al., 1992; Kong et al., 1992). The brace proteins form an ATP-dependent, structure-specific DNA-binding protein complex that recognizes the primer–template junction (Figure 1B). The sliding clamp is then loaded onto the DNA adjacent to the brace proteins, forming a ring that encircles the duplex DNA behind the primer–template junction. A dimer of the β protein forms a ring around the DNA, whereas a trimer of the 45 protein forms a ring that is likely to be structurally very similar to the β dimer. The 45 monomer is two-thirds the mass of the β monomer, indicating how proteins with different primary sequences, mass, and multimeric state can nevertheless adopt similar tertiary structures and perform similar functions. The clamp stimulates the ATPase activity of the brace proteins. When bound to the primer–template, the brace and clamp load the DNA polymerase onto the DNA. Addition of all four dNTPs allows efficient and processive DNA synthesis for many thousands of nucleotides without displacement of the proteins from the template.

Interestingly, the lagging strand is synthesized by the same apparatus. Because the DNA is synthesized in a discontinuous manner, there is a constant need for primers to initiate the short (~1000 nt long) Okazaki fragments. Thus, an additional protein, the DNA primase (DnaG or 61 protein), is needed to form short RNA primers that are then elongated by the polymerase, aided by its brace and

Table 1. Functions of DNA Replication Fork Proteins

E. coli	Phage T4	SV40/ Human	Functions
DnaB	41	T antigen	DNA helicase; stimulates priming on single-stranded DNA
DnaC	59	T antigen	Allows loading of helicase and primase on to SSB-coated DNA (primosome assembly)
SSB	32	RPA	Single-stranded DNA binding; stimulates DNA polymerase; facilitates helicase loading
γ Complex (γδ.δ′χψ)	44/62	RFC	DNA-dependent ATPase; primer–template binding; stimulates DNA polymerase
τ	43	?	Dimerization of holoenzyme
β	45	PCNA	Stimulates DNA polymerase; stimulates DNA-dependent ATPase
Pol III core (αθε)	43	Pol δ[a]	DNA polymerase; 3′→5′ exonuclease
DnaG	61	Primase Pol α[b]	Primase
Ligase	T4 ligase	Ligase 1	Ligation of DNA
DNA pol 1	43	FEN-1 or ME-1	Nuclease for removal of RNA primers
RNase H	RNase H	RNase H1	Nuclease for removal of RNA primers

[a] DNA polymerase δ has been shown to function in SV40 DNA replication; an essential DNA polymerase, DNA polymerase ε, has not yet been assigned a specific function in DNA replication.
[b] The human DNA polymerase α and primase activities function as a multiprotein complex to synthesize RNA–DNA primers.

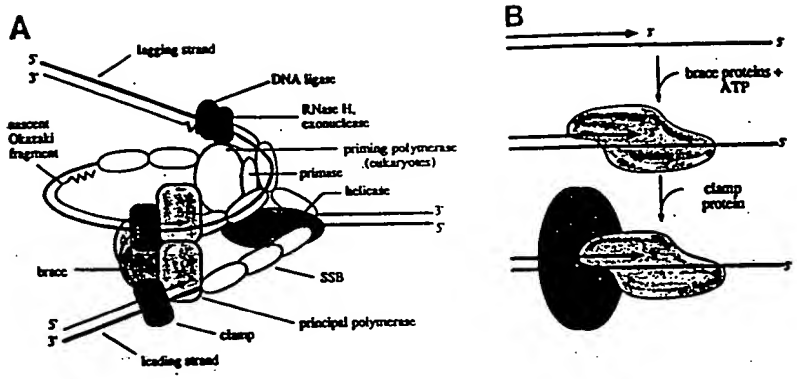


Figure 1. Proposed Generic Model for the Prokaryote and Eukaryote DNA Replication Fork

(A) As suggested by B. Alberts, a dimeric DNA polymerase coordinately replicates the two DNA strands.

(B) Primer recognition by the brace and sliding clamp proteins. The brace proteins (γ complex of E. coli, 44/62 proteins from T4 phage, and RFC from eukaryotes) bind to the primer–template junction in an ATP-dependent manner. The sliding clamp (β protein from E. coli, 45 protein from T4, and PCNA from eukaryotes) binds to the brace and DNA in a ring-like structure. The brace–clamp complex then attracts the DNA polymerase.

clamp (Stukenberg et al., 1991, 1994 [this issue of *Cell*]; Hacker and Alberts, 1994a, 1994b). Both leading- and lagging-strand DNA replication occur coordinately, facilitated by dimerization of the two polymerase complexes. The E. coli τ protein mediates pol III dimerization, whereas the 43 polymerase probably exists as a dimer. Another difference between these machines is that the E. coli pol III holoenzyme ($\alpha\varepsilon\theta.\gamma\delta\delta'\chi\psi.\tau.\beta$) forms a stable protein complex without DNA, whereas the T4 holoenzyme (43.44/62.45) is assembled on the template.

A paradox arises when considering how these machines perform highly processive and continuous DNA synthesis on the leading strand yet, in another context, cycle on and off the lagging strand during discontinuous synthesis of Okazaki fragments. Recent studies of the phage and bacterial machines have suggested a solution to this paradox by demonstrating that the polymerase machines are smart enough to detect the different modes of DNA replication (Hacker and Alberts, 1994a, 1994b; Stukenberg et al., 1994). Both groups have suggested that the polymerase recognizes a key difference between leading- and lagging-strand DNA replication; the polymerase synthesizing a nascent Okazaki fragment will encounter the 5′ end of the previous Okazaki fragment, whereas the leading-strand polymerase is free of such roadblocks.

Stukenberg et al. (1994) have studied the fate of individual proteins when the polymerase machine runs into duplex DNA, taking advantage of the availability of large amounts of the pol III components and the ability to label individual protein components specifically. Design of these experiments was aided by knowledge of the three-dimensional structure of the β clamp. The surface of the clamp that interacts with the polymerase core was modified by addition of a protein kinase recognition site. By following the rate of phosphorylation of this site by excess protein kinase, association and dissociation of the clamp from the polymerase core were measured. This technique of site-specific protein "footprinting" is one that should become popular in studies of protein–protein interaction. When the polymerase core and brace proteins (γ complex) were traveling along the SSB-coated single-stranded DNA template (mimicked by removing one of the four dNTPs and tricking the polymerase into believing it is productively replicating DNA), both were firmly attached to the β clamp and the template. When, on the other hand, the holoenzyme was moving along the template DNA and encountered a DNA duplex (as in the case of a downstream Okazaki fragment), the β clamp dissociated from the other components and remained associated with the newly synthesized DNA (Figure 2A). Dissociation of the polymerase and brace from the template DNA allowed the polymerase

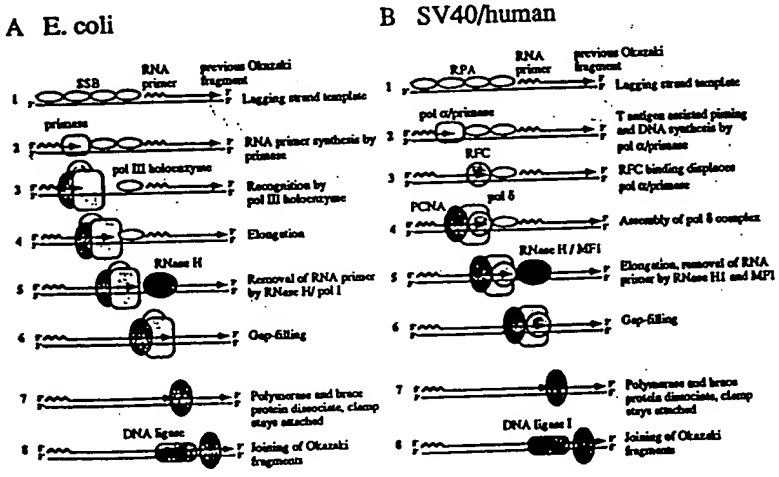


Figure 2. Steps in the Replication of the Lagging Strand

See text for details. Adapted from Stuckenberg et al. (1994) (A) and from Waga and Stillman (1994) (B). Step 7 in (B) is hypothetical, based on E. coli and phage T4 studies.

to cycle onto a different primer–template, facilitated by another clamp and brace attached to the new primer. These studies demonstrated that the holoenzyme has the capacity to detect a different molecular environment and modify its activity, depending on the nature of the template DNA.

Parallel conclusions can be drawn from the studies of Hacker and Alberts (1994a, 1994b). Based on their thorough understanding of the phage T4 replication enzymes, they skillfully engineered hairpin duplexes into the DNA templates downstream of the primer that acted as reversible barriers to the DNA polymerase. In these studies, the kinetics of the dissociation of the T4 holoenzyme was measured under two conditions. When the polymerase was traversing a 32 protein–coated single-stranded DNA template (again mimicked by omitting one dNTP), the replicating polymerase had a half-life on the template of about 150 s, enough time to synthesize more than 75,000 nucleotides (Hacker and Alberts, 1994b). In contrast, when the holoenzyme encountered a duplex DNA hairpin in the path of the moving polymerase, the half-life of the bound polymerase was only 1 s (Hacker and Alberts, 1994a). Thus, the polymerase was smart enough to recognize the new environment and respond appropriately. One would presume that a doltish polymerase would bang headlong into a duplex brick wall in much the same way as a steam locomotive would be unable to recognize a deadend track.

The E. coli pol III holoenzyme replicated the single-stranded template DNA right up to the 5′ end of the downstream duplex DNA located in its path, thereby providing a substrate for the DNA ligase to join the newly synthesized strand to the DNA strand that blocked progression of the polymerase. Normally, during replication of the lagging-strand DNA template, an RNA primer is removed either by an RNase H activity or by the 5′ to 3′ exonuclease activity of E. coli DNA pol I. If the pol III holoenzyme can load onto an RNA primer and replicate the lagging-strand template all the way to the 5′ end of the previously synthesized Okazaki fragment, then only one DNA polymerase molecule might be necessary for lagging-strand DNA replication (Figure 2A). This raises the question of how removal of the RNA primer is coordinated with replication and the role, if any, for DNA pol I and its 5′ to 3′ exonuclease activity in lagging-strand DNA replication. In the case of phage T4, the phage-encoded RNase H removes the primer RNA, and one suspects that the phage-encoded polymerase is the only polymerase required to replicate T4 DNA. The polymerase dissociation mechanism also raises the question of what determines the length of an Okazaki fragment. One strong possibility consistent with these results is that the frequency of initiation on RNA primers that are laid down by primase is the major determinant of Okazaki fragment size (Zechner et al., 1992).

The phage and bacterial DNA polymerase holoenzymes are likely to recognize additional traffic or road blocks. For example, the T4 holoenzyme can pass a transcribing RNA polymerase without dissociating itself or the RNA polymerase from the template (Liu et al., 1993). In another context, the replication machinery might recognize damaged DNA (perhaps bound by damage recognition proteins) and stall or dissociate from the template until the damage is corrected by the DNA repair apparatus. If dissociation occurs, one attractive possibility is that the clamp protein that remains associated with the replicated DNA might itself act as a signal for the repair machinery.

The Eukaryote Fork: Similarities and Differences

An understanding of the protein machinery that functions at the replication fork during duplication of cellular DNA in eukaryotes has not been forthcoming owing to the lack of a suitable biochemical system. Studies on the replication of simian virus 40 (SV40) DNA, however, have provided valuable insight into the eukaryotic replication machine and have revealed many functional similarities and a few differences with the prokaryote machines.

One caveat to relying on SV40 to provide insight into the cellular machinery is that the virus-encoded T antigen performs many functions required for replicating the virus genome. First, it is an initiator protein that functions to begin DNA replication at the SV40 origin; second, it is a DNA helicase that unwinds the DNA ahead of the polymerizing machinery; third, it acts as a primosome-loading protein (Stillman, 1994). Nevertheless, SV40 must rely heavily on the cellular DNA replication apparatus for replication of its own DNA, and biochemical studies over the last 10 years have led to the identification and characterization of these proteins (see Table 1).

In eukaryotes, three DNA ploymerases (α, δ, and ε) have been identified with catalytic subunits having primary amino acid sequence similarity to each other and to the T4 phage 43 protein (Kornberg and Baker, 1992). Unlike its prokaryote cousins, the eukaryotic DNA primase forms a permanent complex with a DNA polymerase (α). Contrary to what was thought for many years, the role of the polymerase α–primase complex appears to be solely for the purpose of providing RNA–DNA primers for initiation of leading-strand synthesis and for initiation of each Okazaki fragment synthesis during lagging-strand replication (Figure 2B) (Nethanel et al., 1988; Tsurimoto et al., 1990; Waga and Stillman, 1994).

PCNA and replication factor C (RFC) are essential subunits of DNA polymarase δ and, under certain conditions, also stimulate the activity of DNA polymerase ε (Lee et al., 1991; Podust and Hubscher, 1993; Waga and Stillman, 1994). In almost all respects, PCNA and RFC function in a very similar way to the β or 45 sliding clamp proteins and the γ complex or 44/62 brace proteins, respectively (Figure 1B) (Tsurimoto et al., 1990; Lee et al., 1991). There is a clear role for both proteins in the replication of the leading and lagging strands during SV40 DNA synthesis in vitro. For SV40 DNA replication in vitro, DNA polymerase δ is responsible for replication of the leading strand and for completion of the lagging strand (see Figure 2B; Waga and Stillman, 1994). Herein lies a significant difference between the prokaryote and eukaryote DNA replication mechanisms (Figure 2). It appears that a dimer of one polymerase replicates both strands at the prokaryotic fork, including the entire Okazaki fragment. In contrast, for SV40 DNA replication, there is a DNA polymerase switch from α to δ during initiation at the replication origin and for synthesis of each Okazaki fragment. Thus, polymerase

δ seems to play a role like the pol III ε/β protein, and it may well be that polymerase δ also cycles between PCNA clamps on the lagging strand, triggered by duplex DNA downstream of the polymerase.

Maturation of the Okazaki fragments in eukaryotes requires a 5′ to 3′ exonuclease (FEN-1 or MF-1) and RNase H1 to remove the RNA from the 5′ end of the Okazaki fragments and DNA ligase I to covalently join the DNA (Turchi et al., 1994; Waga and Stillman, 1994). If PCNA remains at the junction between adjacent Okazaki fragments like the β clamp of E. coli, then it is possible that PCNA may facilitate the activity of one or more of these maturation proteins.

DNA polymerase ε is essential for the viability of the yeast Saccharomyces cerevisiae, but a biochemical role for this polymerase has not been demonstrated in SV40 DNA replication in vitro or for cellular DNA replication. It has been suggested that polymerase ε either functions at the cellular DNA replication fork or, alternatively, plays a role in replication-linked DNA repair pathways that are essential for cell viability. Further work is needed to resolve the function of this enzyme.

Control of Fork Movement

A mechanism that alters the stability of the DNA polymerase holoenzyme on the DNA template could be useful for coordinating DNA replication with other cellular processes. I have already suggested that the holoenzyme may encounter DNA damage or damage-binding proteins and temporarily arrest while DNA repair occurs. Recent studies have demonstrated direct control of the DNA replication function of PCNA by the DNA damage-inducible, p53-regulated, cyclin-dependent kinase inhibitor p21 (Waga et al., 1994). Perhaps the p21 protein, by binding to PCNA directly, tricks the polymerase δ–RFC–PCNA holoenzyme into a state that causes rapid dissociation of the elongating polymerase, thereby arresting DNA replication. If PCNA remained bound to the template, it might facilitate DNA repair at that site. Consistent with this hypothesis, it has been shown that PCNA (and, I suspect, RFC) participates in nucleotide excision repair of ultraviolet-damaged DNA. Thus, the sliding clamp may be a key link between DNA replication and repair.

Regulators such as p21 might be one mechanism to couple DNA metabolism to cell cycle controls. I suspect, however, that we are only at the beginning of a real understanding of the intricacies of these machines and how they are regulated.

References

Capson, T. L., Benkovic, S. J., and Nossal, N. G. (1991). Cell 65, 249–258.

Gogol, E. P., Young, M. C., Kubasek, W. L., Jarvis, T. C., and von Hippel, P. H. (1992). J. Mol. Biol. 224, 395–412.

Hacker, K., and Alberts, B. (1994a). J. Biol. Chem. 269, in press.

Hacker, K., and Alberts, B. (1994b). J. Biol. Chem. 269, in press.

Kong, X.-P., Onrust, R., O'Donnell, M., and Kuriyan, J. (1992). Cell 69, 425–437.

Kornberg, A., and Baker, T. A. (1992). DNA Replication, Second Edition (New York: W. H. Freeman and Company).

Lee, S.-H., Kwong, A. D., Pan, Z.-Q., and Hurwitz, J. (1991). J. Biol. Chem. 266, 594–602.

Liu, B., Wong, M. L., Tinker, R. L., Geiduschek, E. P., and Alberts, B. M. (1993). Nature 366, 33–39.

Munn, M. M., and Alberts, B. M. (1991). J. Biol. Chem. 266, 20034–20044.

Nethanel, T., Reisfeld, S., Dinter-Gottlieb, G., and Kaufmann, G. (1988). J. Virol. 6, 2867–2873.

O'Donnell, M., Onrust, R., Dean, F. B., Chen, M., and Hurwitz, J. (1993). Nucl. Acids Res. 21, 1–3.

Podust, V. N., and Hubscher, U. (1993). Nucl. Acids Res. 21, 841–846.

Stillman, B. (1994). J. Biol. Chem. 269, 7047–7050.

Stukenberg, P. T., Studwell-Vaughan, P. S., and O'Donnell, M. (1991). J. Biol. Chem. 266, 11328–11334.

Stukenberg, P. T., Turner, J., and O'Donnell, M. (1994). Cell 78, this issue.

Tsurimoto, T., Melendy, T., and Stillman, B. (1990). Nature 346, 534–539.

Turchi, J. J., Huang, L., Murante, R. S., Kim, Y., and Bambara, R. A. (1994). Proc. Natl. Acad. Sci. USA 91, 9803–9807.

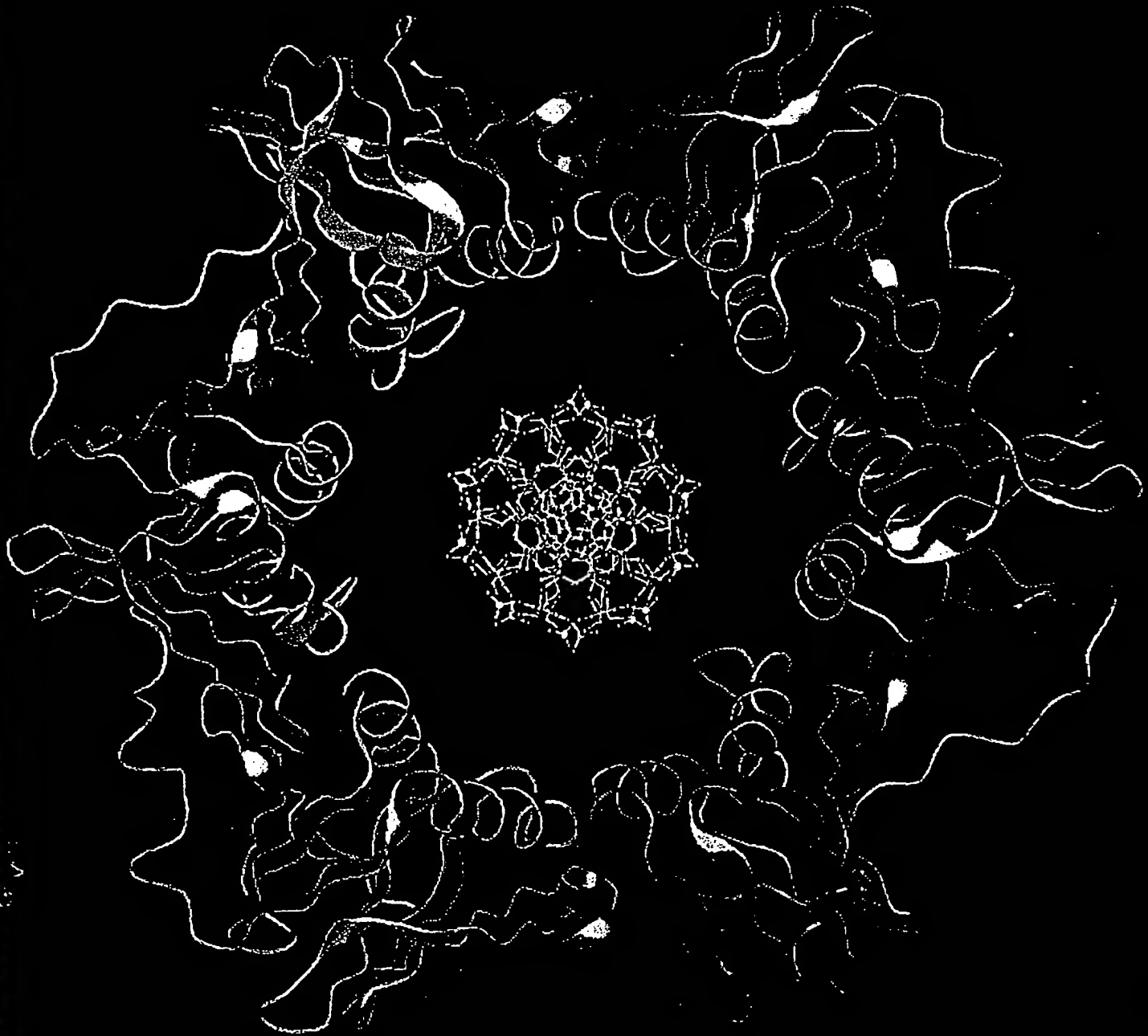
Waga, S., and Stillman, B. (1994). Nature 369, 207–212.

Waga, S., Hannon, G. J., Beach, D., and Stillman, B. (1994). Nature 369, 574–578.

Zechner, E. L., Wu, C. A., and Marians, K. J. (1992). J. Biol. Chem. 267, 4045–4053.

Cell

Volume 69 Number 3 May 1, 1992



Crystal Structure of Phage HK022 Nun Protein Bound to N Boxb RNA and the Role of N and GAP-Associated Proteins and Nucleic Acid Binding

Cell, Vol. 69, 425–437, May [illegible], 1992 Copyright © 1992 by Cell Press

Three-Dimensional Structure of the β Subunit of E. coli DNA Polymerase III Holoenzyme: A Sliding DNA Clamp

Xiang-Peng Kong,* Rene Onrust,† Mike O'Donnell,†‡ and John Kuriyan*§
*Laboratory of Molecular Biophysics
§Howard Hughes Medical Institute
The Rockefeller University
New York, New York 10021
†Microbiology Department
Hearst Research Center
‡Howard Hughes Medical Institute
Cornell University Medical College
New York, New York 10021

Summary

The crystal structure of the β subunit (processivity factor) of DNA polymerase III holoenzyme has been determined at 2.5 Å resolution. A dimer of the β subunit (M_r = 2 × 40.6 kd, 2 × 366 amino acid residues) forms a ring-shaped structure lined by 12 α helices that can encircle duplex DNA. The structure is highly symmetrical, with each monomer containing three domains of identical topology. The charge distribution and orientation of the helices indicate that the molecule functions by forming a tight clamp that can slide on DNA, as shown biochemically. A potential structural relationship is suggested between the β subunit and proliferating cell nuclear antigen (PCNA, the eukaryotic polymerase δ [and ε] processivity factor), and the gene 45 protein of the bacteriophage T4 DNA polymerase.

Introduction

DNA polymerases are enzymes that duplicate the information content of DNA by catalyzing the template-directed polymerization of nucleic acids. A distinction can be made between polymerases that are primarily involved in the replication of chromosomal DNA during cell division and those that normally operate on shorter stretches of template during, for example, the repair of damaged DNA. Polymerases in the latter class are generally nonprocessive, i.e., they polymerize only a few nucleotides before dissociating from the template (Kornberg and Baker, 1991). In contrast, the chromosomal replicative polymerase of Escherichia coli, DNA polymerase III (PolIII) holoenzyme, is distinguished by its ability to perform rapid replication (750 bases per second) of very long stretches of DNA without dissociation (Fay et al., 1981; Burgers and Kornberg, 1982; O'Donnell and Kornberg, 1985; Kornberg and Baker, 1991). This property is conferred upon the enzyme by the presence of associated proteins that clamp the polymerase onto primed DNA, in a process that expends ATP energy. This mechanism, first worked out in detail for E. coli PolIII holoenzyme and the bacteriophage T4 DNA polymerase system, appears to operate analogously in the eukaryotic DNA polymerases δ and ε (Kornberg and Baker, 1991). The three-dimensional structure of one polymerase has been determined by X-ray crystallography, that of the Klenow fragment of E. coli PolI (Ollis et al., 1985). Although this is not a highly processive polymerase, the structure has general relevance for understanding the mechanism of the enzymatic subunits of DNA polymerases. No structural information has yet been available, however, for any of the accessory proteins of the processive polymerases.

Intact PolIII holoenzyme is a complex of at least 10 different protein subunits (α, ε, θ, τ, γ, δ, δ′, χ, ψ, and β) (Maki and Kornberg, 1988). The α subunit performs the catalytic polymerase function, and the ε subunit is the 3′-5′ exonuclease. A three-subunit core polymerase subassembly of the holoenzyme, containing α, ε, and θ, is unable to act processively on its own, although it can fill in short single-stranded regions. The highly processive character of the holoenzyme can be reconstituted upon mixing the core polymerase with both the β subunit and the five-protein γ complex (γ, δ, δ′, χ, and ψ) (Wickner, 1976; O'Donnell, 1987; Maki and Kornberg, 1988). The reconstitution of the processive polymerase proceeds in two distinct stages. In the first stage, the γ complex hydrolyzes ATP to transfer the β subunit to the primed template. In the second stage, the core polymerase assembles with the β subunit on DNA to form the processive polymerase. Thus, it is the β subunit that confers the remarkable processivity onto the core polymerase. Study of the minimal number of subunits required to assemble the processive polymerase showed that only the γ and δ subunits of the γ complex are needed to transfer β from solution to the primed template (O'Donnell and Studwell, 1990). Two subunits of the core polymerase, α and ε (as an αε complex), are needed for processive polymerization (Studwell and O'Donnell, 1990). Once on DNA, the β subunit confers complete processivity onto the αε polymerase, even upon subsequent removal of the γ complex (Stukenberg et al., 1991).

Once the γ complex has performed the operation of clamping the β subunit onto DNA, the β subunit is very strongly bound in that it cannot be easily separated from circular DNA. It has, however, been shown to slide freely along duplex DNA, consistent with its role as a clamp that tethers the polymerase core to the template and moves along with the polymerase during replication (Stukenberg et al., 1991). Experiments using restriction endonucleases have revealed that if circular DNA is cut after the β subunit is clamped on, the β subunit completely separates from DNA by sliding to the site of the break and falling off (Stukenberg et al., 1991). These and related experiments in the same study suggest that, in contrast to site-specific DNA-binding proteins such as transcription factors or nucleases, which make specific hydrogen-bonding or other stabilizing interactions with DNA, the β subunit is bound to DNA mainly by virtue of its topology rather than by stabilizing interactions: it was proposed that the β subunit might form a closed ring, and that one role of the γ complex might be to open and then close the ring around DNA,

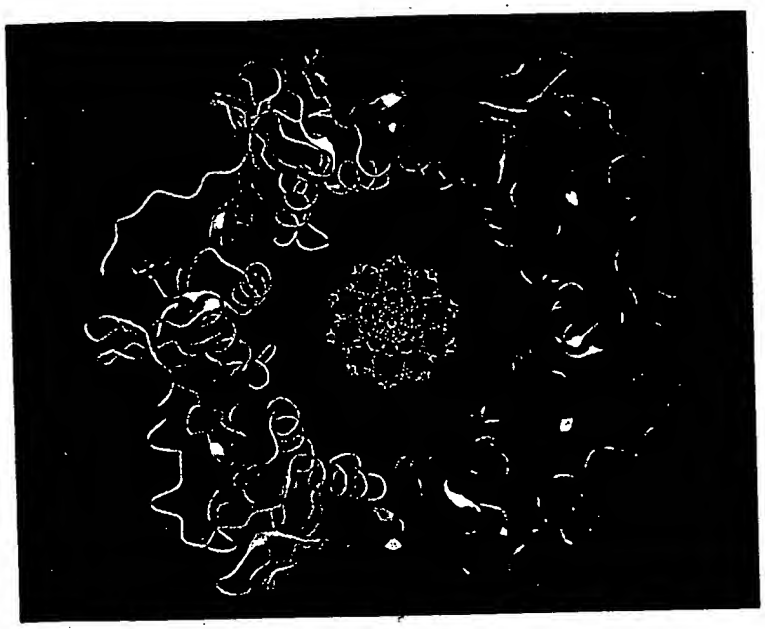
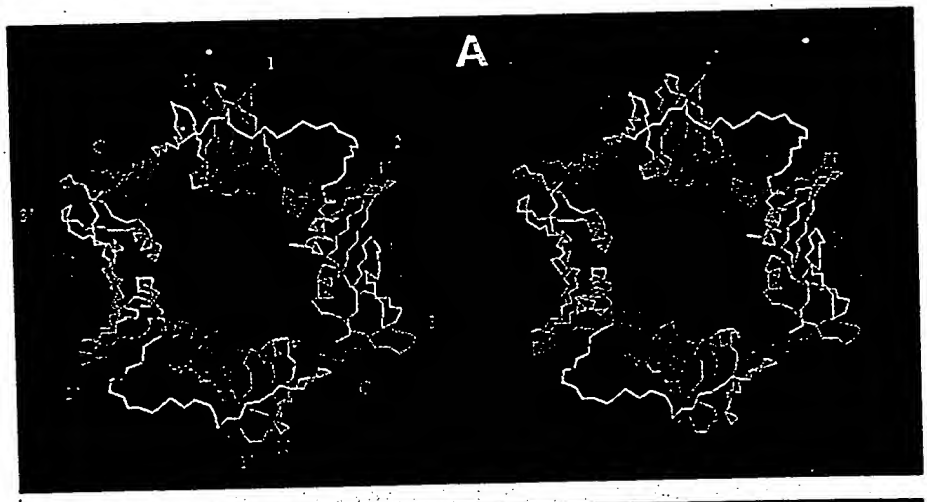
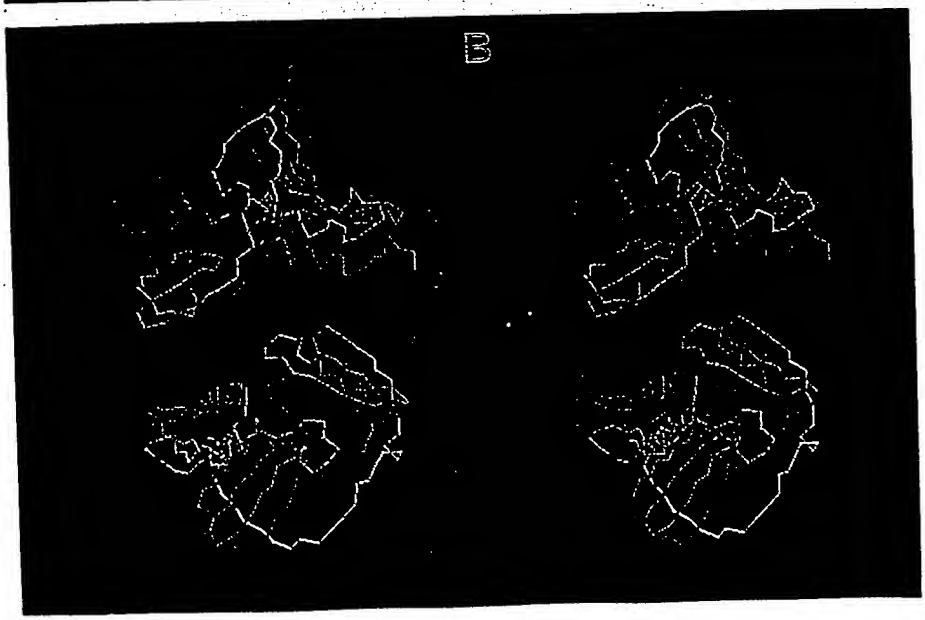


Figure 1. Ribbon Representation of the Polypeptide Chain of a β Subunit Dimer, Looking Down the 2-Fold Axis of the Ring

The α helices are shown as spirals and the β sheets as flat ribbons. The two monomers are colored yellow and red. A standard model of B-form DNA (Saenger, 1984) is in the middle of the structure, represented in stick form with phosphorus, oxygen, nitrogen, and carbon atoms colored yellow, red, blue, and green, respectively. The DNA structure is hypothetical, and is placed in the geometric center of the β subunit ring with the helix axis aligned along the 2-fold rotation axis of the ring. Figures 1, 2, 3, 6, and 8 were generated using the program QUANTA (Polygen Corp.). Refined atomic coordinates and X-ray structure factors are being deposited in the Protein Databank.





effectively trapping the β subunit on DNA (Stukenberg et al., 1991).

As a step toward a detailed understanding of the molecular basis of the processive properties of PolIII holoenzyme, we have crystalized the β subunit and determined its three-dimensional structure by X-ray diffraction at 2.5 Å resolution. In pleasing congruence with previous speculation, we find that the structure is indeed that of a closed ring, with overall shape similar to that of a donut or toroid. In this paper we discuss the consequences of this structure for the function of the β subunit, and suggest a possible structural relationship between this protein and its functional equivalent in the eukaryotic polymerase δ (and ε) replicases, the processivity factor PCNA (proliferating cell nuclear antigen), and the bacteriophage T4 DNA polymerase gene 45 protein. This represents the first atomic resolution view of any of the accessory proteins of processive polymerases involved in DNA replication, and complements the previously reported structure of the polymerase I Klenow fragment (Ollis et al., 1985).

Results and Discussion

Architecture of the β Subunit Dimer

The β subunit forms a head-to-tail dimer in the crystal, consistent with previous observations that the isolated protein is a dimer in solution (Johanson and McHenry, 1980). Representations of the polypeptide backbone of the dimer are shown in Figures 1 and 2, and a space-filling representation of all atoms is in Figure 3. The overall structure is that of a star-shaped ring, of approximate diameter 80 Å, with a hole of diameter ~35 Å in the middle (Figure 1). The 2-fold dimer axis is perpendicular to the face of the ring, the thickness of which is about that of one full turn of duplex B-form DNA (~34 Å).

The starlike shape of the ring is due to an unexpected feature of the structure: it is internally symmetric, with each monomer of the β subunit consisting of three structural domains of identical chain topology and very similar three-dimensional structure. Each domain is roughly 2-fold symmetric in its architecture, with an outer layer of two β sheets providing a scaffold that supports two α helices. Replication of this motif around a circle results in a rigid molecule with 12 α helices lining the inner surface of the ring, and with 6 β sheets forming the outer surface (Figure 1). Surprisingly, this structure is reminiscent of the ring-shaped pentameric assembly of the B subunits of cholera toxin-related heat-labile enterotoxin from E. coli (Sixma et al., 1991). The toxin structure has 5 α helices that line the inner surface of the ring with antiparallel β sheets forming the outer surface. The hole in the ring is plugged by an extended polypeptide strand from the A subunit. The hole in the toxin structure (diameter 11 Å) is much smaller than that observed in the β subunit dimer, and the detailed topology of each of the B subunits of the toxin is different from that of the β subunit modules.

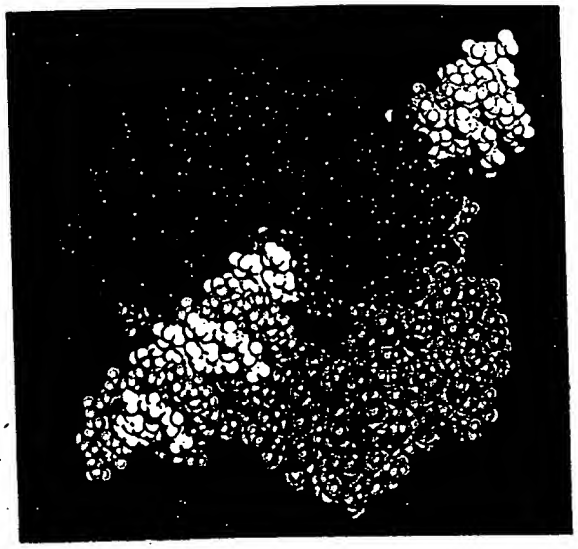


Figure 3. Space-Filling Model of the β Subunit Dimer with B-Form DNA

One monomer is colored red and the other yellow. The radius of the spheres corresponds to the van der Waals radius of the corresponding atom. Hydrogen atoms are not explicitly displayed, but manifest themselves as increased radii for atoms that they are bonded to. The hypothetical model of B-form DNA is as in Figures 1 and 6, and is shown with one strand colored white and the other green. The double helix passes through the hole in the β subunit dimer with no steric repulsions.

The secondary structure and chain topology of a domain is particularly simple and is shared by all three domains in one monomer: two adjacent antiparallel α helices are flanked by two 4-stranded antiparallel β sheets (see schematic diagram in Figure 4A). One of the β sheets forms the commonly found "greek key" motif (Branden and Tooze, 1991). The other β sheet contains the N- and C-termini of the chain. If an imaginary connection is drawn between the termini, this β sheet also forms a greek key, and a striking 2-fold symmetry in the chain topology is revealed (Figure 4A). Although there is insufficient sequence similarity to draw conclusions about the evolution of this fold, we note that one domain can be generated by duplication of a βαββ motif (Figure 4A). The chain topology diagram also reveals the simple principle underlying the architecture of the entire ring. The two outer strands of β sheets in one domain form hydrogen-bonding interactions with corresponding strands in two adjacent domains, contin-

Figure 2. C_α Connectivity of the β Subunit Dimer

The stereo diagrams are colored based on the sequence number of the residues, with the colors smoothly varying in the order green, light blue, purple, red, and yellow, with increasing sequence number within each monomer, from the N-terminus to the C-terminus. The dimer interfaces are between the domains colored green/blue and red/yellow. (A) The 2-fold axis is approximately perpendicular to the page. The approximate locations of the chain termini are marked with N and C. The domains are numbered 1, 2, 3 and 1′, 2′, 3′ in the two monomers. (B) Edge view of the ring.

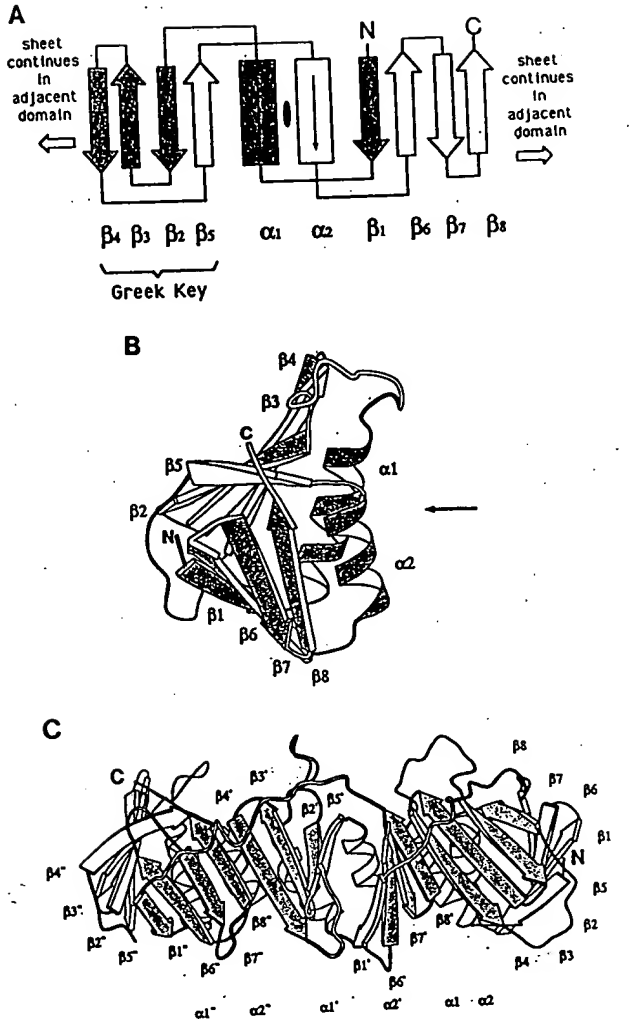


Figure 4. Secondary and Tertiary Structure of the β Subunit

The secondary structure elements were defined using the program DSSP (Kabsch and Sander, 1983).

(A) Schematic diagram of the secondary structure of one domain of the β subunit. α helices are shown as rectangles and β sheets as arrows. The topology is 2-fold symmetric about the point indicated by the ellipse. The structure can be generated by duplication of a βαββββ unit, shaded in gray.

(B) Ribbon diagram (Priestle, 1988) of a domain. The approximate 2-fold symmetry axis is indicated by the arrow. The secondary structural elements are labeled as in (A).

(C) Ribbon diagram of one monomer. The β sheets that are continued across domain boundaries are shaded gray. These constitute two of the six β sheets in the dimer. The other two are the corresponding ones in the other molecule, and two others are formed by continuing the sheets (shown unshaded) across the molecular boundaries. Note the presence of three protruding loops on the top half of the molecule. The secondary structural elements are labeled as in (A), with the unprimed, singly primed, and doubly primed labels referring to the first, second, and third domain, respectively.

ued around the circle (Figure 4C). No distinctions are apparent between such β sheet extensions across internal domain boundaries as opposed to intermolecular contacts, i.e., the two dimer interfaces also form continuous antiparallel β sheets. These interactions lead to a completely closed circle with six "seamless" β sheets on the outer surface (Figures 1, 3, and 4).

Each domain consists of about 110 residues, and forms a compact and well-folded structure (Figure 4B). The 2-fold symmetry apparent in the topology diagram manifests itself as a very approximate 2-fold axis between the two helices (Figure 4B). Despite the simple architecture, each domain is quite clearly an independent folding unit, with a well-defined hydrophobic core consisting of about 20 residues. The symmetry axis relating the two molecules in the dimer is noncrystallographic, i.e., the two monomers are packed into different crystal environments and are crystallographically independent. The transformation that optimally superimposes the two molecules is a 180° rotation about an axis perpendicular to the plane of the ring, and results in a root mean square (rms) deviation of 0.40 Å between equivalent C_α positions (not including residues in the loops formed by residues 18–28 and 210–213, and the last three residues at the C-terminus). This axis is inclined by 12° with respect to the b-axis of the $P2_1$ crystal form, and therefore the holes of translationally related dimers do not line up to generate a linear tunnel in the crystal. Rotations about the dimer axis also superimpose the three different domains that constitute each monomer, and this basic structural unit repeats after every 60° rotation about the axis. Although the amino acid sequences of each domain are quite different, superimposition of the C_α positions of the three domains reveals that approximately 80% of these positions can be considered to be structurally analogous (Figure 5). The rms deviation in the positions of these C_α positions is 1.2 Å.

Potential Mode of Interaction with DNA

In the replication of a chromosome, the initial clamping of a β subunit dimer on DNA occurs at a primer terminus which is RNA/DNA hybrid duplex, presumably A-form (Saenger, 1984; Kornberg and Baker, 1991). Subsequent to this event, the β subunit can freely move along duplex DNA, presumably B-form (Stukenberg et al., 1991). The interactions of the β subunit with both the A and B forms of the double helix are therefore of interest. A-form DNA is similar to the B form in terms of the width of its cross section, and the local direction of the phosphate backbone with respect to the helix axis is very similar in both forms (Saenger, 1984). All subsequent analysis is focused on just these two gross features of the double helix, and the results are applicable to both the A and B forms.

Although the structure has been determined in the absence of DNA, several obvious features indicate that the protein is designed to wrap around the double helix with a minimum of locally specific interactions. The high symmetry of the structure is well suited to interact with the cylindrically symmetric DNA duplex, and the hole in the middle of the ring (of diameter ~35 Å, not including extended sidechains) is large enough to easily accommodate either the A or B forms of DNA (diameter ~25 Å) with no steric repulsion. Insertion of a model of either form of DNA into the ring results in a precise relationship between the common tilt in the orientations of all 12 α helices and the tilt of the phosphate backbone (see below). Finally,

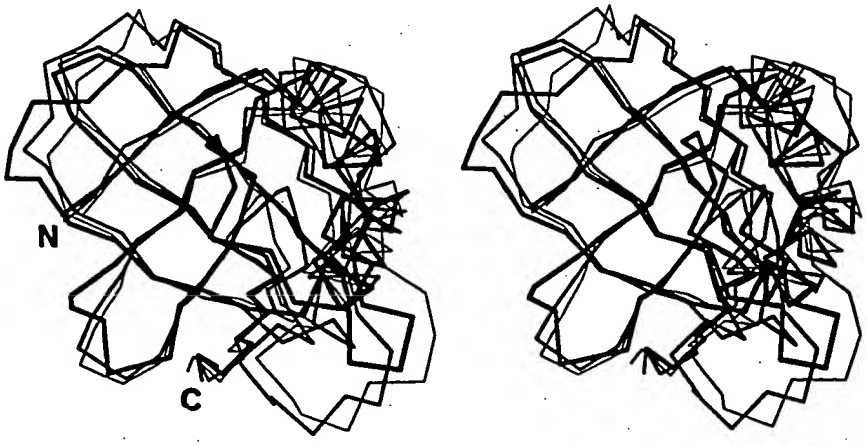


Figure 5. Structural Similarity between Three Domains of the Monomer

The C_α connectivity of each monomer is displayed in a stereo diagram after the three are optimally superimposed. The domains were considered in pairs, and a least-squares algorithm was used to align the structures (QUANTA, Polygen Corp.). This process was started by matching the central portion of one of the β sheets, then increasing the list of atoms included in the matching list and repeating the least-squares optimization. The process was repeated until the list included approximately 80% of the C_α atoms in each domain, with an rms deviation of 1.2 Å between corresponding C_α atoms. The remaining atoms are in the loops and in one of the helices, and are judged to have somewhat altered structure in the three domains. Domains 1 and 3 have very similar structure for both helices. In domain 2, one of the helices (α2) is displaced by about half a helical turn along the helix axis. This does not significantly alter its orientation with respect to the proposed model of DNA.

although the protein is strongly negatively charged, calculation of the electrostatic field generated by the molecule reveals a focusing of positive electrostatic field in the center of the ring, precisely where the negatively charged phosphate backbone of DNA is expected to be.

One consequence of the symmetrical arrangement of the six domains is that each of the 12 α helices has a similar tilt with respect to the axis of the ring. Assuming that duplex DNA passes through the middle of the ring, we generated a model of standard B-form DNA and placed it in the center of the β subunit (Figures 1, 3, and 6). The simple assumption that the duplex is perpendicular to the plane of the ring results in an intriguing relationship between the axes of all 12 helices and the phosphate backbone lining the major and minor grooves of DNA (Figure 6). The axis of each helix is almost precisely perpendicular to the local direction of the phosphate backbone, i.e., the helices span the major and minor grooves. This feature seems designed to prevent entry of the protein into either groove, and should facilitate rapid motion along the duplex. An additional consequence of aligning the DNA perpendicular to the ring is that each helix interacts with a different phase of DNA: if one helix spans the major groove, the one directly across the ring spans the minor groove. This is likely to lead to a damping out of the variation in interaction energy with the phosphate backbone as the protein moves across the grooves of DNA. The α helices are maintained in this precise orientation due to packing interactions with each other and with the underlying β sheet. Despite the intrinsic curvature of the β sheets, the strands, by and large, run in directions parallel to the helices, and therefore perpendicular to the DNA backbone in this model (Figures 2 and 4).

When viewed down the double helix, the phosphate backbone of DNA is 10-fold and 11-fold symmetric in projection for the A and B forms of DNA, respectively (such a projection is shown for the B form in Figure 1). Thus, there is no specific correspondence between the number of helices (12) that line the hole of β subunit and repetitive periods in DNA (10 or 11). Rather, the important features of the structure are the circular symmetry and the conservation of the helix–DNA backbone interaction. Given the dimensions of the ring and the tilt of the helices, the 12 helices pack against each other with no additional space remaining.

A dimer of β subunit contains 38 aspartate, 58 glutamate, 24 lysine, 50 arginine, and 14 histidine residues. The protein thus has a net charge of −22 if all histidines are assumed neutral, and −15 if half of them are charged. This negative charge is consistent with the inability of the β subunit to bind DNA without ATP activation by the γ complex. The electrostatic charge is not, however, uniformly distributed over the protein. We have found it useful to visualize the effects of the asymmetric charge distribution by calculating the electrostatic field generated by the protein, using a continuum electrostatic model that treats the protein as a low dielectric medium with embedded charges, immersed in a high dielectric solvent (water) of variable ionic strength (Gilson et al., 1988). While a quantitative analysis of charge effects would require a careful consideration of approximations introduced in this model, we focus solely on qualitative features that are reproducibly obtained with various calculational parameters (see legend to Figure 7).

Two important qualitative features consistently emerge from analysis of the computed electrostatic field. The outer edge and both faces of the ring are in regions of strongly negative electrostatic potential (Figures 7A and 7B). How-

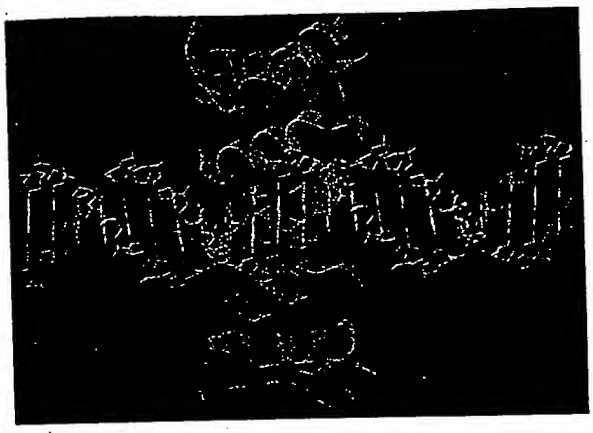


Figure 6. The Helices of the β Subunit Are Perpendicular to the Phosphate Backbone of the DNA

The 12 α helices of the β subunit dimer and a hypothetical model of B-form DNA are shown. The relative orientation of the double helix and the β subunit is exactly as in Figures 1 and 3. For clarity, the α helices are shown schematically in a ribbon representation, with the brighter ones being closer to the viewer. The central helix in front of the DNA is perpendicular to the phosphate backbone and spans the major groove of the double helix, while the one furthest behind it spans the minor groove. Each of the 12 helices has similar disposition with respect to the DNA backbone, but faces a different combination of the major and minor grooves. In this projection, the major and minor grooves are superimposed, with the major groove in the center of the figure being nearer the viewer.

ever, the surface of the hole has strongly positive electrostatic potential, and would be expected to interact favorably with the negatively charged backbone of DNA. Thus, the β subunit dimer can be described as a ring of negative charge surrounding a positively charged core (Figure 7). This focusing of positive charge is probably necessary to stabilize the dimer around DNA, as the dimer interface is itself mainly electrostatic in nature (see discussion below) and is unlikely to withstand repulsive interactions with DNA.

The other important feature is that the two faces of the ring are quite dissimilar in their properties, due to the head-to-tail dimer formation (Figure 7B). Although both are negatively charged, the negative electrostatic potential is clearly more extended on one face than the other. The other face has six prominent loops that extend away from the ring and appear to be well suited for interactions with another protein subunit (Figure 4C). We presume that only one of these faces interacts with the αε polymerase, which is known to bind strongly to the β subunit (Stukenberg et al., 1991). Although the faces of the β subunit are asymmetric, the symmetry of duplex DNA makes it unlikely that one orientation will be preferred over the other. It is likely that the γ complex interacts with the primer template junction in a specific orientation, and may thereby correctly orient the β subunit with respect to the primer terminus for productive interaction with the αε polymerase.

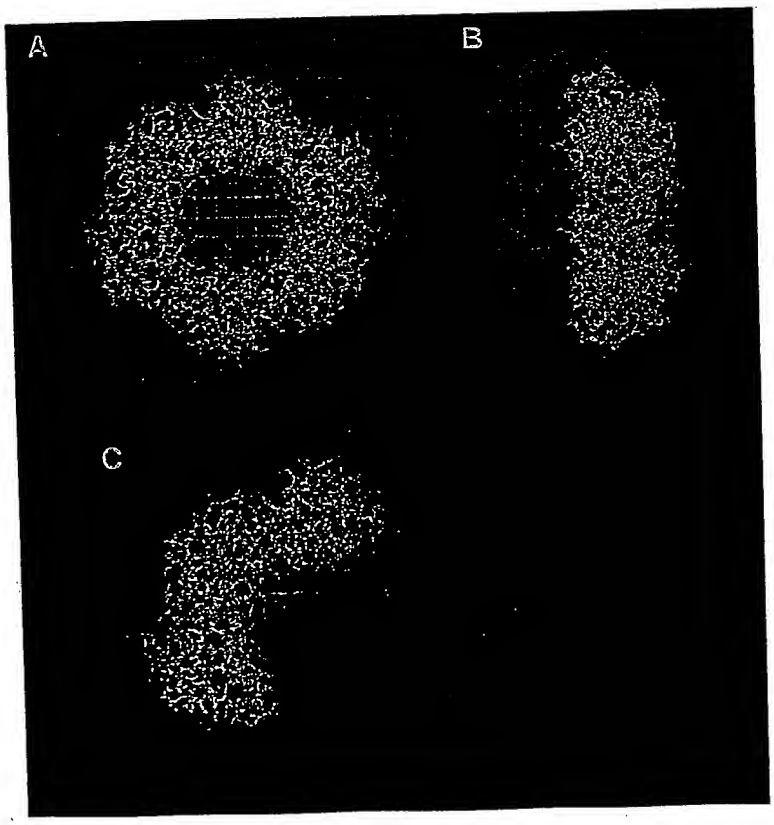


Figure 7. Electrostatic Potential Maps for the β Subunit Dimer and an Isolated Monomer

The maps are calculated using the programs DELPHI (Gilson et al., 1988) and INSIGHTII (Biosym Technologies). Lys and Arg residues have a single positive charge localized on the terminal nitrogen atoms of the sidechains. Asp and Glu residues have a single negative charge, localized on the terminal oxygen atoms of the sidechains. His sidechains have a 1/2 positive charge each. All other atoms in the molecule are considered neutral. Qualitatively similar results are obtained upon changing the charges on the protein to make all histidines neutral. The calculation was done assuming a uniform dielectric of 80 for the solvent and 2 for the protein interior. The ionic strength was set to zero. The red and blue mesh contours represent negative electrostatic potential (energy of $-2.5\ k_BT/e$) and positive electrostatic field (energy of $+2.5\ k_BT/e$), respectively. k_BT is the product of the Boltzmann constant and the temperature, and e is the charge of the electron. Two orthogonal views of the electrostatic potential of the dimer are shown in (A) and (B), and (C) is the potential for an isolated monomer.

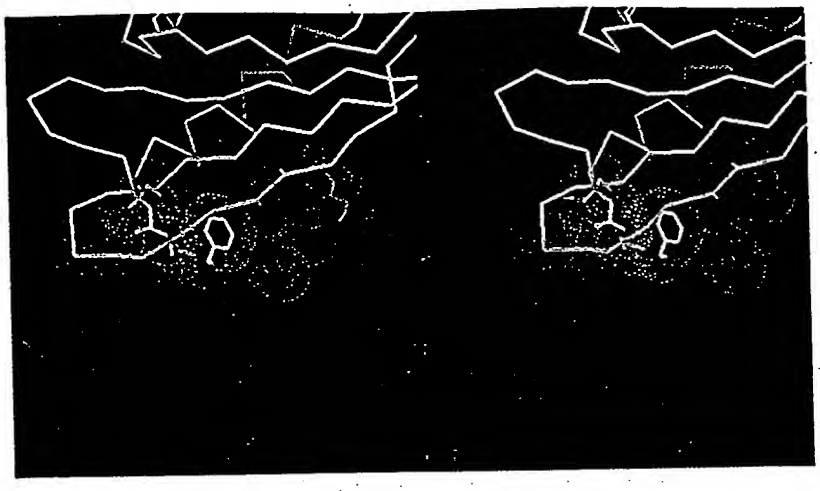


Figure 8. Stereo Diagram of the Dimer Interface

Bonds linking C_α atoms in one monomer are yellow, and those in the other are green. Note the continuation of the β sheet across the interface. Ten amino acid sidechains are shown at the interface, with the dot surfaces representing the van der Waals spheres of the atoms. Four of these are hydrophobic residues and are colored white (Phe-106 and Ile-108 from one monomer and Leu-273 and Ile-272 from the other). Arg-103, Arg-105, and Lys-74 are contributed by one monomer and are colored blue. Glu-300, Glu-303, and Glu-304 are from the other monomer and are colored red. For clarity, Arg-96 and Glu-303 are not shown.

The spacing between the hypothetical phosphate backbone of DNA and the protein sidechains is such that direct contacts are not likely to be made. The distance of closest approach between the terminal atoms of fully extended arginine residues lining the hole and the phosphate groups is expected to be no less than 3.5 Å. This suggests that water molecules will play an important role in mediating the protein–DNA interactions, which will also increase the ability of the protein to move along DNA without becoming attracted to any particular region. This lack of specificity makes it unlikely that useful crystals of a β subunit–DNA complex would ever be obtained. Molecular dynamics simulations of a solvated model of β subunit and DNA may be a useful approach toward understanding the details of this interaction.

The Dimer Interface and the Formation of the Head-to-Tail Dimer

The nature of the dimer interface is relevant to an understanding of the role played by the γ complex and ATP in assembling the β subunit on duplex DNA. The main feature of the interface is a continuation, across the molecular boundary, of β sheet structure (Figures 2, 8, and 10). This appears indistinguishable from β sheet continuation at interdomain boundaries within a monomer, and contributes at least four strong hydrogen bonds at each of the two interfaces. Relatively little surface area is buried upon formation of the β subunit dimer. The exposed surface area calculated using a water-sized probe of radius 1.5 Å (Lee and Richards, 1971) is 33,218 Å^2 for two isolated monomers, and decreases by only 8% (to 30,525 Å^2) upon dimerization.

In addition to the hydrogen bonds contributed by the β sheet, further stabilization at the dimer interface is provided by two distinct sets of interactions between amino acid sidechains from neighboring domains (Figure 8). At the center of the interface, the sidechains of Phe-106 and Ile-278 from one monomer pack against Ile-272 and Leu-273 from the other and form a small hydrophobic core. Surrounding these residues are six potential intermolecular ion pairs: Lys-74 and Glu-298′ (closest observed distance between charged groups of 2.9 Å, with the prime indicating a residue in the other monomer), Lys-74 and Glu-300′ (2.5 Å), Arg-96 and Glu-300′ (4.7 Å), Arg-103 and Glu-304′ (3.0 Å), Arg-105 and Glu-301′ (3.3 Å), and Arg-105 and Glu-303′ (3.0 Å). Two of the ion pairs (Arg-96–Glu-300′ and Arg-103–Glu-304′) involve charged groups that are both inaccessible to solvent, as determined using a water-sized probe (Lee and Richards, 1971), which is expected to lead to particularly strong ionic interactions. A feature of these interactions is that all the positively charged residues are contributed by one monomer, with the other one contributing the negatively charged residues. The computed electrostatic potential for an isolated monomer reflects this charge asymmetry, with the N- and C-terminal parts of the monomer being in regions of positive and negative electrostatic potential, respectively (Figure 7C). This extensive electrostatic complementarity is unique to the dimer interface, and is not observed at the interdomain boundaries within a monomer.

The dimer interface is thus seen to have a number of specific and potentially strong interactions, and a plausible explanation for the ATP requirement for β subunit assembly on DNA is that energy is required to break the interfacial hydrogen bonds and buried ion pairs. Nevertheless, the relatively small interaction surface suggests that a monomer–dimer equilibrium is possible in solution. In this regard, it is interesting to contrast the observed head-to-tail dimer (with the N-terminal domain of one monomer interacting with the C-terminal domain of the other) with a

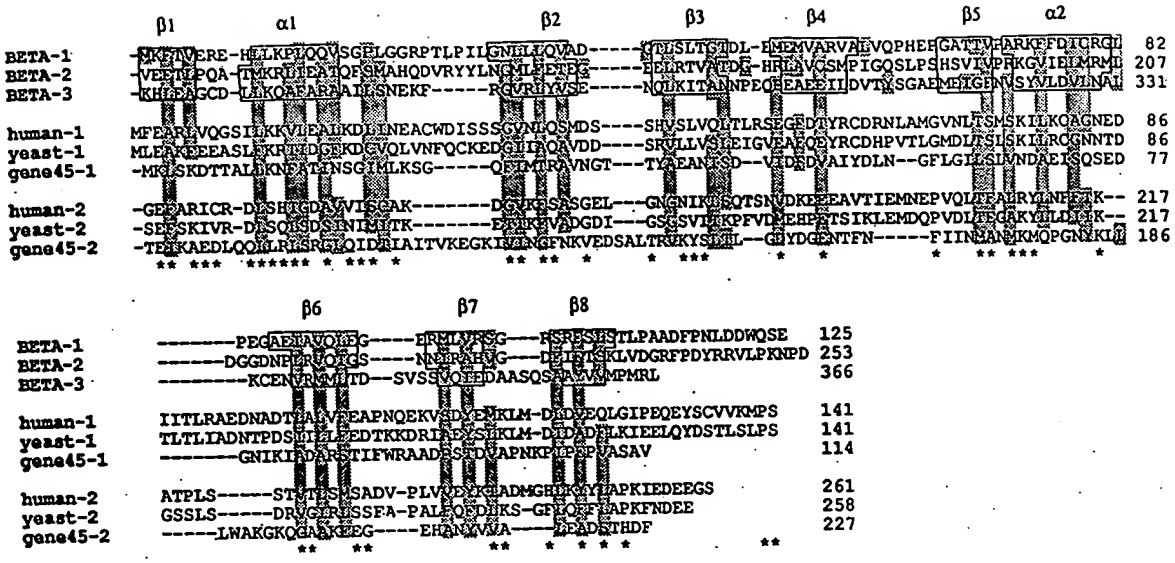


Figure 9. Alignment of the Sequences of the Domains of the β Subunit with Human PCNA, Yeast PCNA, and Gene 45 Protein

The three domains of the β subunit are labeled BETA-1, BETA-2, and BETA-3, and are aligned based on three-dimensional structural correspondence alone. The human, yeast, and gene 45 sequences have been split into two domains labeled 1 and 2. The secondary structural elements within the β subunit domains are boxed and labeled. The meanings of the shaded bars for the β subunit sequences are different from those for the PCNA and gene 45 sequences. Within the β subunit, the bars indicate all the residues that are completely buried, as judged by solvent accessibility calculations using a water-sized probe and the intact dimer (Lee and Richards, 1971). No criterion of sequence similarity was used in the selection of these residues. For the PCNA and gene 45 sequences, the shaded bars indicate residues that are deemed to be similar in property to those that are buried in at least two of the β subunit domains. The residues Asp, Glu, Gln, Asn, His, Arg, and Lys have been excluded from the gray bars in PCNA and gene 45 unless a charged residue is present in the corresponding β subunit. The asterisks underneath the sequences indicate positions where at least two of the PCNA and gene 45 domains have sequence identity with amino acids in the β subunit domains. The numbers to the right are the last residue numbers in each row.

head-to-head assembly. The latter would have identical faces on both sides of the ring and is not forbidden from considerations of chain topology alone. Model building shows that by rotating one of the monomers, a structure can be constructed which maintains similar β sheet and α helix packing at the interface, but in which the N- and C-terminal domains interact with the corresponding domains from the other monomer (this would correspond to rotating one of the monomers by 180° about a vertical axis in Figure 2B). Such an arrangement would, however, lead to unfavorable electrostatic interactions, as the N- and C-terminal regions of the monomers are positively and negatively charged, respectively (Figure 7C). Electrostatic complementarity is therefore likely to play an important role in favoring an initial head-to-tail association that could then lead to the formation of the complete interface and the burial of ion pairs.

Sequence Comparison between β Subunit Domains, PCNA, and Gene 45 Protein

The internal symmetry and domain structure of the β subunit was not suspected earlier because the three domains share very little sequence identity. Pairwise sequence alignments, based on the three-dimensional structure, result in only 16%, 9%, and 9% amino acid identity between domains 1 and 2, 2 and 3, and 1 and 3, respectively (Figure 9). This is well below the threshold of 20%–25% identity required for the predication of similar three-dimensional structure (Sander and Schneider, 1991). Likewise, conventional sequence-matching algorithms fail to detect any relationship between the sequences of the three domains. However, knowledge of the three-dimensional structure makes possible an unambiguous internal sequence alignment, based simply on structural correspondence, and this reveals two conserved features (Figure 9). First, there are 17 positions in each domain that are completely buried and where only Phe, Leu, Ile, Met, Val, or Ala is present. These constitute the conserved hydrophobic cores of the domains. There are another 10 or so buried positions where charged residues are not tolerated (with the exception of residues involved in dimer formation; see below). Second, although the protein is overall negative in charge, the two α helices in each domain have net positive charge (the 12 helices have a total of 22 Arg and Lys residues and 8 Glu and Asp residues).

The three domains presumably diverged from an ancestral module by gene duplication events. That the present level of sequence identity is so low probably reflects the fact that the structural constraints on the domain are fairly loose compared to the geometric requirements at, for example, an enzyme active site or the recognition site of a site-specific transcription factor. Even in systems that do have such geometric requirements, it appears that large numbers of different sequences can lead to the same functional fold as demonstrated by random mutagenesis experiments on λ repressor (Lim and Sauer, 1991). The di-

vergence in sequence between the three domains of the β subunit is not surprising in this context.

Given the lack of significant sequence similarity between the three domains of the β subunit, it seems unlikely that conventional sequence-matching schemes will be successful in recognizing protein sequences that adopt this three-dimensional fold. We have, instead, focused our attention on proteins that are known to be functionally similar to the β subunit, and looked to see if they are likely to have the same overall architecture. These proteins are the various "processivity factors" of other replicative polymerases (Kornberg and Baker, 1991), the best characterized of which are the bacteriophage T4 DNA polymerase system (Nossal and Alberts, 1984) and the yeast and mammalian chromosomal replication polymerase, Pol δ (Tsurimoto and Stillman, 1990; Burgers, 1991; Lee et al., 1991a, 1991b). For highly processive DNA replication, each of these polymerases requires a set of ATP-dependent accessory factors that appear to act to tether the core polymerase enzyme onto DNA, in a manner analogous to the γ complex and β subunit of E. coli PolIII holoenzyme. The accessory factor that has functional correspondence with the β subunit is the gene 45 protein for the T4 DNA polymerase system, and PCNA for eukaryotic Pol δ (Kornberg and Baker, 1991). T4 gene 45 protein and PCNA are approximately two-thirds the size of the β subunit (T4 gene 45 protein: 227 residues; Saccharomyces cerevisiae PCNA: 258 residues; human PCNA: 261 residues). Given their smaller size and knowing the three-domain structure of the β subunit, we reasoned that these proteins may be composed of two domains each, and thereby function as trimeric molecules that assemble six domains around DNA in sets of two per monomer. There is no obvious amino acid sequence homology between the β subunit and T4 gene 45 protein or any PCNA. However, measurements of the molecular weight of T4 gene 45 protein and S. cerevisiae PCNA in solution indicate that they are likely to be trimeric (Jarvis et al., 1989; Bauer and Burgers, 1988).

The sequences of T4 gene 45 protein and of yeast and human PCNA were compared with those of the three domains of the β subunit, using the following criteria. Insertions and deletions were allowed only between the secondary structural elements of the β subunit, and the hydrophobic core and other buried residues of β were matched with hydrophobic or neutral polar residues in the other sequences. The yeast and human sequences are closely related (30% identity) (Almendral et al., 1987; Bauer and Burgers, 1990), and these conditions led to a plausible, albeit weak, alignment for both PCNA sequences with the β subunit (Figure 9). A similar alignment is obtained for gene 45 protein with the β subunit (Figure 9). This alignment differs from one previously reported between yeast PCNA and T4 gene 45 protein, which included several very long insertions and deletions (Tsurimoto and Stillman, 1990).

The resulting alignment between the β subunit domains and other sequences (two domains each) is shown in Figure 9, which also indicates the buried residues in the β subunit. The alignment preserves in PCNA the hydrophobic core of the β subunit, and also reveals a suggestive level of sequence identity between PCNA and the domains of the β subunit. At 37 positions out of 110, there are at least two amino acid identities between the domains of the β subunit and PCNA (Figure 9). Some of these sequence identities are intriguing. For example, sheet 4 in the third domain of the β subunit has two buried Glu residues that are involved in dimer formation. These are conserved in corresponding positions in the PCNA sequences. Two of the positively charged residues involved in ion pairing at the dimer interface (Lys-74 and Arg-96) have corresponding Lys or Arg residues at the same position or one amino acid removed in the PCNA sequences. Eight positions in the domains of the β subunit have at least one buried aromatic residue. Five of these have corresponding aromatic residues in one of the PCNA sequences. Finally, although both PCNA sequences have a net excess of negatively charged residues, the regions corresponding to the helices have net positive charge, qualitatively similar to what is observed in the β subunit. Normalized to 12 α helices, the yeast sequences have 18 positive and 12 negative residues on the helices (the human sequence has an additional negative charge). Similar conclusions can be drawn from the alignment of gene 45 protein with the β subunit (Figure 9). As in the β subunit and PCNA, although the overall charge of the gene 45 protein is negative, the net charge for the α helices is positive (15 positively charged and 9 negatively charged when normalized to 12 helices). Ion pairs can also be formed in the trimer interface of the gene 45 protein. The negative charged Asp-49 and Asp-51 (which are aligned one residue away from the buried Glu residues) can be paired with the positively charged Lys-177 and -184, or alternatively, the positively charged Arg-127 and -130 can be paired with Asp-69, Glu-71, and Glu-76.

This alignment suggests that if PCNA and gene 45 form a ring around DNA, then they must do so as trimers, with each monomer of the protein contributing two domains that are roughly similar in architecture to the β subunit domains. However, it must be stressed that the alignment is weak and should only be taken as a hypothesis for further experimental testing. For example, the recently developed profile method of sequence comparison (Bowie et al., 1991), which relies on knowledge of the three-dimensional structure of one of the structures being compared, is unable to detect any significant similarity between β subunit domains and PCNA or gene 45 protein when provided with the β subunit structure. It is, however, able to detect some, but not all, of the internal symmetry of the β subunit. A conclusive understanding of PCNA and gene 45 protein architecture must await determination of the three-dimensional structures.

Conclusion

A satisfying feature of the structure of the β subunit is the beautiful way in which the circular symmetry of the double helix has been reflected in the ring-shaped and highly symmetric structure of the protein clamp. This symmetry manifests itself on a smaller scale than the domain boundaries, and the entire structure can be considered to be generated by replication of a simple βαββ motif 12 times around a

circle, with the β strands and α helices being oriented perpendicular to the direction of the grooves of DNA. This simple and elegant structure carries out a simple function and stands in sharp contrast to the structure of the actual catalytic subunit of the polymerase, as probably exemplified by the Klenow fragment of PolI. That enzyme, which must recognize the information content of DNA and then replicate one strand, has none of the symmetry of DNA and resembles rather the palm of a hand gripping the double helix (Ollis et al., 1985). For processive polymerases involved in DNA replication, it appears that the proper combination of symmetry and asymmetry is a requirement for functionality.

Experimental Procedures

Crystallization

The β subunit was purified to >99% purity as described (Onrust et al., 1991), with the following modifications. Chromatography on AH–Sepharose was performed in place of SP–sephadex, and the chromatofocusing step was replaced by chromatography on a fast flow Q column and a mono-Q column (Pharmacia). The purified β subunit was concentrated to 16 mg/ml, as determined by Bradford assay (Bradford, 1976), by ultrafiltration in a buffer containing 20% glycerol, 20 mM Tris buffer (pH 7.5), and 0.5 mM EDTA. The protein solution in glycerol was used as such for all further crystalization trials, using a sparse matrix method (Jancarick and Kim, 1991). Crystalization was rapid, with many different conditions yielding crystals within a day or two of the first experiments. Optimization of several of these conditions led to three different crystal forms (Forms I, II, and III) that are suitable for high-resolution X-ray diffraction analysis.

All crystals were grown using the hanging drop method (McPherson, 1990). The drops were initially set up by mixing 2 μl of protein and 2 μl of a "reservoir" solution, and equilibrated against the reservoir at room temperature or 4°C. The unit cell constants and space groups of the various crystal forms were determined by a combination of oscillation data collection (see below), single counter diffractometry (using a Rigaku AFC5 diffractometer), and precession photography. Form I crystals are in space group P1 ($a = 86.8$ Å, $b = 73.9$ Å, $c = 65.7$ Å, $\alpha = 75.2°$, $\beta = 86.8°$, $\gamma = 81.8°$) and are grown at 4°C from reservoirs containing 13%–15% isopropanol, 100 mM $CaCl_2$, and 100 mM MES buffer (pH 7.5). Form II crystals are obtained at room temperature (21°C) from reservoirs containing 30% polyethylene glycol (average $M_r = 400$), 100 mM $CaCl_2$, and 100 mM MES (pH 6.5). The space group is P1 ($a = 41.7$ Å, $b = 72.9$ Å, $c = 65.5$ Å, $\alpha = 74.6°$, $\beta = 85.1°$, $\gamma = 82.2°$). Forms I and II are related, with the asymmetric unit in Form I being a doubling of that in Form II. Finally, Form III crystals are obtained at room temperature (21°C) from conditions very similar to those that yield Form II crystals, except that the pH is lowered to 6.0. This crystal form is in space group $P2_1$ ($a = 80.6$ Å, $b = 68.3$ Å, $c = 82.3$ Å, $\beta = 114.2°$).

All three crystal forms grow to large sizes ($0.3 \times 0.4 \times 0.7$ mm³) and diffract strongly to 2 Å resolution. The presence of the rotational symmetry axis in Form III crystals makes them much more suitable for oscillation data collection than the lower-symmetry P1 forms, particularly for the simultaneous measurement of Bijvoet pairs, and thus the structure determination was carried out using this form alone. Based on molecular volume calculations, Forms II and III contain a dimer of the β subunit in the asymmetric unit, while Form I contains two dimers. Assuming this stoichiometry, the volume per unit mass, V_m, as defined by Matthews, is 2.5, 2.3, and 2.6 Å³ for the three forms, respectively, within the range typical for protein crystals (Matthews, 1968).

Data Collection and Structure Determination

X-ray intensity data collection was carried out by the oscillation method (Arndt and Wonacott, 1977), using a Rigaku R-AXIS IIC imaging phosphor area detector, mounted on a Rigaku RU200 rotating anode X-ray generator (Molecular Structure Corp., Houston). Typical crystal-to-detector distances and exposure times were 153 mm and 15 min, respectively, for 2° oscillations. Data processing and reduction were done entirely by software provided by Rigaku (Table 1).

The structure determination was carried out by the multiple isomorphous replacement (MIR) method. The amino acid sequence of the β subunit shows that the molecule contains four cysteines (Ohmori et al., 1984), at least one of which is reactive to N-ethylmaleimide (Johanson et al., 1986). The search for isomorphous heavy atom derivatives was therefore focused on mercury compounds, and two good derivatives were obtained using mercuric chloride and ethyl mercury phosphate. The binding sites for mercury are similar in both cases, with mercuric chloride and ethylmercury phosphate reacting with three and four cysteines per monomer, respectively (Table 2). Although crystals of the native protein have reasonable lifetimes in the X-ray beam, allowing complete data sets to be collected on single crystals, the derivative crystals exhibit much earlier radiation decay. For the ethylmercury phosphate derivative, this was overcome by measuring data at −5°C. The mercuric chloride–treated crystals were not stable at the lower temperature, and a total of four crystals were used for data collection at room temperature (Table 2). For both derivatives, measurement of anomalous differences was optimized by aligning crystals so that the 2-fold rotation axis (b^*) was precisely along the oscillation axis, leading to simultaneous measurement of Bijvoet pairs.

Difference Patterson maps for the mercuric chloride derivative showed a small number of strong peaks (10–16 standard deviations above the mean density value) in the $y = 1/2$ Harker section, and heavy

Table 1. Refined Heavy Atom Positions

Derivative		Relative Occupancy	x	y	z	Derivatized Cysteine	Distance to Sulphur (Å)
$HgCl_2$	1	1.0	0.26	0.40	0.44	79	1.88
	2	1.0	0.41	0.80	0.99	79′	2.14
	3	0.6	0.00	0.86	0.09	180	1.60
	4	0.6	0.84	0.50	0.52	180′	1.69
	5	0.6	0.83	0.39	0.88	333	1.95
	6	0.6	0.00	0.54	0.53	333′	1.51
EMP	1	0.6	0.26	0.40	0.43	79	1.37
	2	0.4	0.41	0.80	0.98	79′	1.68
	3	0.8	0.01	0.86	0.09	180	1.52
	4	0.9	0.84	0.50	0.52	180′	1.81
	5	0.3	0.83	0.39	0.88	333	1.90
	6	0.3	0.00	0.54	0.53	333′	1.59
Sites not present in $HgCl_2$ derivative:							
	7	0.9	0.81	0.38	0.93	260	1.60
	8	1.0	0.02	0.50	0.49	260′	2.16

The positions are fractional coordinates in the unit cell, and the prime on the cysteine number indicates that it is in the second monomer.

Table 2. Statistics for Data and Derivatives

	Native	$HgCl_2$ I	$HgCl_2$ II	EMP
Number of crystals	1	3	1	1
Concentration (mM)		2	2	2
Soaking time (hr)		12	12	6
Resolution (Å)	2.4	2.78	3.0	2.79
Measured reflections	101,590	86,408	42,097	61,458
Unique reflections	30,957	19,660	16,148	19,568
Completion (%)	94.4	93.3	97.3	94.1
R-merge[a] (%)	7.29	7.87	7.41	6.78
Mean isomorphous difference[b] (%)		24.4	25.6	17.8
Phasing power[c]		2.1	2.1	1.4
Mean figure of merit	0.67			

The two $HgCl_2$ data sets are for differently oriented crystals (see text). [a] $\Sigma_h\Sigma_i|I_{hi} - I_h|/\Sigma_h\Sigma_i I_{hi}$, where I_{hi} is the scaled intensity of the i^{th} observation of reflection h, and I_h is the mean value. [b] $\Sigma||F_{PH}| - |F_P||/\Sigma|F_P|$, where F_{PH} and F_P are the scaled derivative and native structure factor amplitudes, respectively. [c] Phasing power: rms heavy-atom structure factor / rms lack of closure.

atom positions were readily determined by manual inspection. Additional sites were found by difference Fourier techniques. Heavy atom parameters were refined and initial phases were calculated using the program HEAVY (Terwilliger and Eisenberg, 1983), including anomalous differences for only those reflections that were fully recorded on one image (71.7% and 70.2% of the data had anomalous measurements included for mercuric chloride and ethylmercury phosphate, respectively). The quality of the anomalous data was judged to be good, since the anomalous difference Patterson map recapitulated the major peaks in the difference Patterson map. For calculating and refining the MIR phases, the mercuric chloride data were divided into two sets. The first set consists of data (including anomalous differences) from three crystals, each of which had been oriented with the b^*-axis parallel to the oscillation axis. The second set consists of data from a crystal mounted in an arbitrary orientation and does not include anomalous differences, and better results were obtained by treating these data separately in the phase calculations. Summary statistics from the heavy atom phasing procedure are given in Table 1.

The MIR phases were further improved by solvent flattening (Wang, 1985) using a program written by W. Kabsch (COMBINE). These phases were then used to generate a 3 Å electron density map. Predominant features of the structure, such as the donut shape, the two-layer architecture with helices in the middle and β sheets outside, and the approximate 6-fold symmetry, were immediately recognized. The program BONES (Jones and Thirup, 1986) was used to generate a skeleton representation of the density, which made evident the fact that the structure is made up of six copies of very similar modules. An additional feature that emerged at this stage was that the 2-fold noncrystallographic symmetry apparent in the mercury positions correlated extremely well with a 2-fold noncrystallographic symmetry in the electron density. This 2-fold axis is only 12° away from the crystallographic 2-fold screw axis, and was therefore not identified in self-rotation functions.

A partial model of poly-Ala chain was built for 80% of the structure fairly readily, using the program FRODO and a data base of protein structures (Jones and Thirup, 1986). During this process, partial atomic models were repeatedly used to improve the phases by phase combination with the MIR and solvent-flattened phases, using COMBINE. Although the noncrystallographic symmetry was never explicitly used in the phase refinement process, it provided a useful check on the accuracy of the phases, as did the unexpected but obvious 3-fold symmetry within each monomer. The topology of the protein fold and

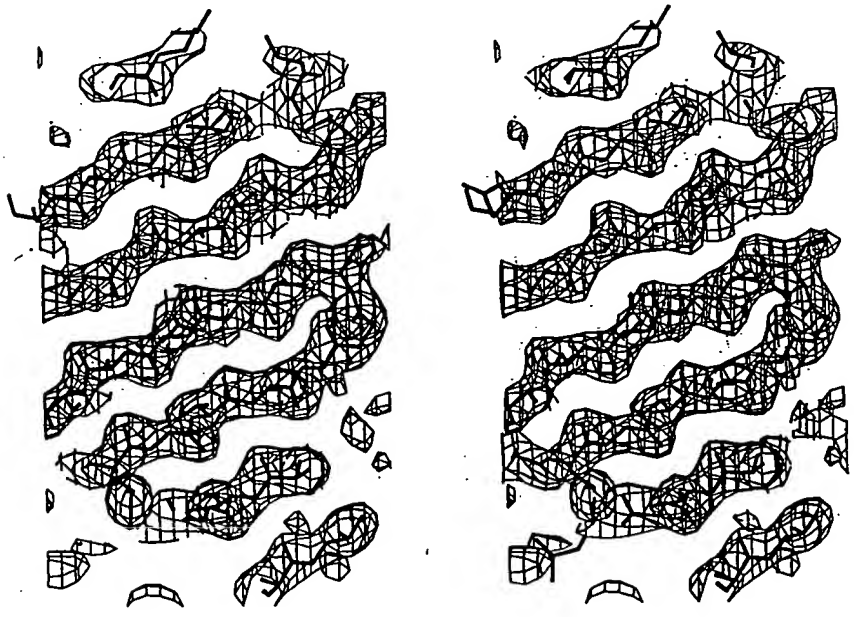


Figure 10. Stereo View of the Electron Density

Electron density at 2.5 Å resolution, at one of the two dimer interfaces of the β subunit. The three β strands at the top are from one molecule and the other four from the second molecule. Electron density contours are displayed at 1 σ above the mean value of the density. Note that the directions of the carbonyl oxygens are apparent as bumps in the backbone density, which makes the direction of the chain unambiguous. The quality of the electron density map apparent here is typical of the rest of the map, which has no breaks in backbone density anywhere in the molecule. The displayed map is a $(2|F_o|-|F_c|)\exp\{i\alpha_c\}$ Fourier synthesis, where F_c and F_o are calculated and observed structure factors, respectively, and α_c is the calculated phase. The model used is the final refined one.

the locations of the N- and C-termini of the monomers were clearly recognized at this stage. A complete model of one monomer with the correct sequence (Ohmori et al., 1984) was readily built (except for a loop of 10 residues that was left as poly-Ala) by assuming that the mercury atoms had bound to cysteine residues, and tracing the sequence out from these points. The second monomer was generated from this model by the noncrystallographic symmetry operation.

This atomic model was refined against 3.0 Å native X-ray data by least-squares refinement with both monomers treated independently, using the program X-PLOR (Brünger, 1988). Two hundred steps of Powell optimization and 40 steps of individual B factor refinement smoothly reduced the R factor from 46.2% to 24.9% with no manual intervention, indicating the accuracy of the initial model. Repeated model building using FRODO and least-squares refinement using X-PLOR were carried out. A total of five rounds of simulated annealing refinement were then carried out, using initial and final temperatures of 1000 K and 300 K, respectively (Brünger et al., 1987; Weis et al., 1990). The resolution of the native data included was extended to 2.5 Å, and 150 well-resolved solvent molecules (interpreted as water) were included. The current model has the entire sequence built in, and has unbroken backbone electron density from the N- to the C-terminus in difference Fourier maps. The R factor is 18.9% (for 27,614 reflections with $|F|>2\sigma(F)$) and the rms deviation from ideal geometry is 0.017 Å for bond lengths and 3.5° for bond angles, with good stereochemistry for the backbone torsion angles. Representative electron density for the final refined model is shown in Figure 10, at the dimer interface. Although the electron density is weak for a few surface sidechains (arginines and lysines in particular), the backbone density is strong throughout the structure, including all surface loops. The orientation of the carbonyl groups is clear for almost all residues (Figure 10).

Acknowledgments

We thank T. S. R. Krishna, Stephen K. Burley, and Gregory A. Petsko for helpful suggestions, and W. Kabsch for the program COMBINE. This work was supported in part by grants from the National Institutes of Health (GM 43094 to J. K. and GM 38839 to M. O'D.). J. K. is a Pew Scholar in the Biomedical Sciences.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 USC Section 1734 solely to indicate this fact.

Received January 17, 1992; revised February 18, 1992.

References

Almendral, J. M., Huebsch, D., Blundell, P. A., Macdonald-Brown, H., and Bravo, R. (1987). Cloning and sequence of the human nuclear protein cyclin: homology with DNA binding proteins. Proc. Natl. Acad. Sci. USA *84*, 1575–1579.

Arndt, V. W., and Wonacott, A. J. (1977). The Rotation Method in Crystallography (Amsterdam: North-Holland).

Bauer, G. A., and Burgers, P. M. J. (1988). The yeast analog of mammalian cyclin/proliferating-cell nuclear antigen interacts with mammalian DNA polymerase δ. Proc. Natl. Acad. Sci. USA *85*, 7506–7510.

Bauer, G. A., and Burgers, P. M. J. (1990). Molecular cloning, structure and expression of the yeast proliferating cell nuclear antigen gene. Nucl. Acids Res. *18*, 261–265.

Bowie, J. U., Lüthy, R., and Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. Science *253*, 164–170.

Bradford, M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. Anal. Biochem. *72*, 248–254.

Branden, C., and Tooze, J. (1991). Introduction to Protein Structure (New York: Garland).

Brünger, A. T. (1988). X-PLOR (Version 2.2) Manual. Howard Hughes Medical Institute and Dept. of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut.

Brünger, A. T., Kuriyan, J., and Karplus, M. (1987). Crystallographic R-factor refinement by molecular dynamics. Science *235*, 458–460.

Burgers, P. M. J. (1991). *Saccharomyces cerevisiae* replication factor C. II. Formation and activity of complexes with proliferating cell nuclear antigen and with DNA polymerases δ and ε. J. Biol. Chem. *266*, 22698–22706.

Burgers, P. M. J., and Kornberg, A. (1982). ATP activation of DNA polymerase III holoenzyme of *Escherichia coli*. J. Biol. Chem. *257*, 11468–11473.

Fay, P. J., Johanson, K. O., McHenry, C. S., and Bambara, R. A. (1981). Size classes of products synthesized processively by DNA polymerase III and DNA polymerase III holoenzyme of *Escherichia coli*. J. Biol. Chem. *256*, 976–983.

Gilson, M. K., Sharp, K. A., and Honig, B. H. (1988). Calculating the electrostatic potential of molecules in solution: method and error assessment. J. Comp. Chem. *9(4)*, 327–335.

Jancarick, J., and Kim, S.-H. (1991). Sparse matrix sampling: a screening method for crystallization of proteins. J. Appl. Crystallogr. *24*, 409–411.

Jarvis, T. C., Paul, L. S., and von Hippel, P. H. (1989). Structural and enzymatic studies of the T4 DNA replication system. I. Physical characterization of the polymerase accessory protein complex. J. Biol. Chem. *264*, 12709–12716.

Johanson, K. O., and McHenry, C. S. (1980). Purification and characterization of the β subunit of DNA polymerase III holoenzyme of *Escherichia coli*. J. Biol. Chem. *255*, 10984–10990.

Johanson, K. O., Haynes, T. E., and McHenry, C. S. (1986). Chemical characterization and purification of the β subunit of the DNA polymerase III holoenzyme from an overproducing strain. J. Biol. Chem. *261*, 11460–11465.

Jones, T. A., and Thirup, S. (1986). Using known substructures in protein model building and crystallography. EMBO J. *5*, 819–822.

Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers *22*, 2577–2637.

Kornberg, A., and Baker, T. A. (1991). DNA Replication (New York: W. H. Freeman).

Lee, B. K., and Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. J. Mol. Biol. *55*, 379–400.

Lee, S.-H., Kwong, A. D., Pan, Z.-H., and Hurwitz, J. (1991a). Studies on the activator 1 protein complex, an accessory factor for proliferating cell nuclear antigen-dependent DNA polymerase δ. J. Biol. Chem. *266*, 594–602.

Lee, S.-H., Pan, Z.-Q., Kwong, A. D., Burgers, P. M. J., and Hurwitz, J. (1991b). Synthesis of DNA by DNA polymerase ε *in vitro*. J. Biol. Chem. *266*, 22707–22717.

Lim, W. A., and Sauer, R. T. (1991). The role of internal packing interactions in determining the structure and stability of a protein. J. Mol. Biol. *219*, 359–376.

Maki, S., and Kornberg, A. (1988). DNA polymerase III holoenzyme of *Escherichia coli*. III. Distinctive processive polymerases reconstituted from purified subunits. J. Biol. Chem. *263*, 6561–6569.

Matthews, B. W. (1968). Solvent content of protein crystals. J. Mol. Biol. *33*, 491–497.

McPherson, A. (1990). Current approaches to macromolecular crystallization. Eur. J. Biochem. *189*, 1–23.

Nossal, N. G., and Alberts, B. M. (1984). Mechanism of DNA replication catalyzed by purified T4 replication proteins. In Bacteriophage T4 (Washington, D.C.: American Society for Microbiology).

O'Donnell, M. E. (1987). Accessory proteins bind a primed template and mediate rapid cycling of DNA polymerase III holoenzyme from *Escherichia coli*. J. Biol. Chem. *262*, 16558–16565.

O'Donnell, M. E., and Kornberg, A. (1985). Complete replication of templates by *Escherichia coli* DNA polymerase III holoenzyme. J. Biol. Chem. *260*, 12884–12889.

O'Donnell, M., and Studwell, P. S. (1990). Total reconstitution of DNA polymerase III holoenzyme reveals dual accessory protein clamps. J.

Biol. Chem. *265*, 1179–1187.

Ohmori, H., Kimura, M., Nagata, T., and Sakakibara, Y. (1984). Structural analysis of the *dnaA* and *dnaN* genes of *Escherichia coli*. Gene *28*, 159–170.

Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G., and Steitz, T. A. (1985). Structure of large fragment of *Escherichia coli* polymerase I complexed with dTMP. Nature *313*, 762–766.

Onrust, R., Stukenberg, P. T., and O'Donnell, M. (1991). Analysis of the ATPase subassembly which initiates processive DNA synthesis by DNA polymerase. J. Biol. Chem. *266*, 21681–21686.

Priestle, J. P. (1988). RIBBON—stereo cartoon drawing program for proteins. J. Appl. Crystallogr. *21*, 572–576.

Saenger, W. (1984). Principles of Nucleic Acid Structure (New York: Springer-Verlag).

Sander, C., and Schneider, R. (1991). Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins *9*, 56–68.

Sixma, T. K., Pronk, S. E., Kalk, K. H., Wartna, E. S., van Zanten, B. A. M., Witholt, B., and Hol, W. G. J. (1991). Crystal structure of the cholera toxin-related heat-labile enterotoxin from *E. coli*. Nature *351*, 371–377.

Studwell, P. S., and O'Donnell, M. (1990). Processive replication is contingent on the exonuclease subunit of DNA polymerase III holoenzyme. J. Biol. Chem. *265*, 1171–1178.

Stukenberg, P. T., Studwell-Vaughan, P. S., and O'Donnell, M. (1991). Mechanism of the sliding β-clamp of DNA polymerase III holoenzyme. J. Biol. Chem. *266*, 11328–11334.

Terwilliger, T. C., and Eisenberg, D. (1983). Unbiased three-dimensional refinement of heavy-atom parameters by correlation of origin-removed Patterson functions. Acta Crystallogr. *A39*, 813–817.

Tsurimoto, T., and Stillman, B. (1990). Functions of replication factor C and proliferating-cell nuclear antigen: functional similarity of DNA polymerase accessory proteins from human cells and bacteriophage T4. Proc. Natl. Acad. Sci. USA *87*, 1023–1027.

Wang, B. C. (1985). Resolution of phase ambiguity in macromolecular crystallography. Meth. Enzymol. *115*, 90–112.

Weis, W. I., Brünger, A. T., Skehel, J. J., and Wiley, D. C. (1990). Refinement of the influenza virus hemagglutinin by simulated annealing. J. Mol. Biol. *212*, 737–761.

Wickner, S. (1976). Mechanism of DNA elongation catalysed by *Escherichia coli* DNA polymerase III, dna Z protein, and DNA elongation factors I and III. Proc. Natl. Acad. Sci. USA *73*, 3511–3515.

Cell, Vol. 79, 1233–1243, December 30, 1994, Copyright © 1994 by Cell Press

Crystal Structure of the Eukaryotic DNA Polymerase Processivity Factor PCNA

Talluru S. R. Krishna,* Xiang-Peng Kong,*
Sonja Gary,† Peter M. Burgers,† and John Kuriyan*‡
*Laboratories of Molecular Biophysics
‡Howard Hughes Medical Institute
The Rockefeller University
New York, New York 10021
†Department of Biochemistry and Molecular Biophysics
Washington University School of Medicine
St. Louis, Missouri 63110

Summary

The crystal structure of the processivity factor required by eukaryotic DNA polymerase δ, proliferating cell nuclear antigen (PCNA) from S. cerevisiae, has been determined at 2.3 Å resolution. Three PCNA molecules, each containing two topologically identical domains, are tightly associated to form a closed ring. The dimensions and electrostatic properties of the ring suggest that PCNA encircles duplex DNA, providing a DNA-bound platform for the attachment of the polymerase. The trimeric PCNA ring is strikingly similar to the dimeric ring formed by the β subunit (processivity factor) of E. coli DNA polymerase III holoenzyme, with which it shares no significant sequence identity. This structural correspondence further substantiates the mechanistic connection between eukaryotic and prokaryotic DNA replication that has been suggested on biochemical grounds.

Introduction

The rapid replication of chromosomes relies on DNA polymerases that initiate replication in response to regulatory signals, achieve high processivity without dissociation from the template, and then disengage rapidly and restart replication elsewhere as needed (reviewed by Kornberg and Baker, 1991; Kuriyan and O'Donnell, 1993; Stillman, 1994). Mechanistic studies on the chromosomal replicases of Escherichia coli, bacteriophage T4, and eukaryotes have shown that these multisubunit polymerase complexes include a distinct processivity factor, variously referred to as a sliding clamp or a DNA-tracking protein, that is required for nondissociative DNA replication. The clamp is first loaded onto primed DNA by auxiliary proteins in an ATP-dependent process, and subsequently the polymerase machinery attaches to the clamp and starts processive DNA replication. The utilization of a processivity factor that is distinct from the catalytic subunits allows for both rapidity and control. During chromosomal replication, additional DNA clamps are loaded onto each newly primed template on the lagging strand, and the catalytic machinery can switch from one clamp to the next after an Okazaki fragment is completed (Stukenberg et al., 1994). On the other hand, sequestration of the clamp or inactivation of the proteins needed to load the clamp onto DNA might provide a way for the cell to control the progression of DNA replication.

Mammalian proliferating cell nuclear antigen (PCNA), so named because of its initial discovery as a cell cycle–dependent antigen (Miyachi et al., 1978), is the processivity factor or sliding clamp for DNA polymerase δ (Pol δ) and is essential for the replication of SV40 and BPV viral DNA by this enzyme (Prelich et al., 1987a, 1987b; Wold et al., 1988; Tan et al., 1986; Bravo et al., 1987; Muller et al., 1994). In yeast, the POL30 gene encoding PCNA is essential for cell growth, showing a requirement for PCNA in chromosomal DNA replication (Bauer and Burgers, 1988). PCNA is also a processivity factor for Pol ε, another essential eukaryotic DNA polymerase (Morrison et al., 1990; Burgers, 1991; Lee et al., 1991b; Podust and Hübscher, 1993). In addition, PCNA and Pol δ also play essential roles in the repair of damaged DNA (Shivji et al., 1992; Zeng et al., 1994). In nontransformed cells, PCNA is found in complexes with a variety of cyclins, cyclin-dependent kinases, and the p21 protein inhibitor of cyclin-dependent kinases, also known as Waf1 or Cip1 (Xiong et al., 1993; Zhang et al., 1993). This suggests a potential role for PCNA in the response to cellular regulatory signals, particularly since the p21 protein can bind directly to PCNA and thus inhibit DNA replication (Waga et al., 1994; Flores-Rozas et al., 1994).

Reconstitution of total replication of DNA containing viral origin sequences reveals that PCNA is required for the complete replication of both the leading and the lagging strands by Pol δ (Waga and Stillman, 1994; Muller et al., 1994). DNA polymerase α and primase lay down short stretches of RNA and DNA on the template to start DNA replication. Recognition of the resulting template–primer junction by the multipolypeptide replication factor C (RFC), also known as activator 1, results in the loading of PCNA onto the DNA and the inhibition of further synthesis by Pol α (Tsurimoto and Stillman, 1990; Lee et al., 1991a). Pol δ then binds to PCNA and carries out processive DNA replication, either continuously for the leading strand or discontinuously for the lagging strand (Okazaki fragment synthesis). Elimination of any of RFC, PCNA, or Pol δ prevents leading strand synthesis and results in the generation of incomplete Okazaki fragments (Waga and Stillman, 1994). In these viral replication systems, there is no apparent in vitro requirement for Pol ε. However, during chromosomal DNA replication, which is undoubtedly more complex, Pol ε may also play a role in lagging strand synthesis in a PCNA-dependent manner (Burgers, 1991; Podust and Hübscher, 1993).

The broad outline of the mechanism of processive DNA replication in eukaryotes that is now emerging has elements in common with the mechanisms of the well-studied Pol III holoenzyme of E. coli and the T4 DNA polymerase (Stillman, 1994). The functional equivalents of PCNA in these prokaryotic replicases are the β subunit of Pol III and the gene 45 protein of T4 phage (Cha and Alberts, 1988; Kornberg and Baker, 1991; Kuriyan and O'Donnell,

1993) (here referred to as β subunit and gene 45 protein, respectively). Although the processivity factors function by forming a strong association with DNA, the interaction is topological rather than chemical, and these molecules cannot normally bind to DNA without assistance from additional factors (Stukenberg et al., 1991; Tinker et al., 1994; Kong et al., 1992). Thus, another required component of these systems is the protein complex (γ complex in E. coli, the products of genes 44/62 in T4 phage, and the RFC complex in eukaryotes) that recognizes and binds to the duplex–single-strand junctions that are initiation points for replication and then function to load the DNA clamps onto DNA in an ATP-dependent manner (Kornberg and Baker, 1991). Some sequence similarity has been noted between the subunits of the γ complex, T4 phage genes 44/62, and the RFC complex (O'Donnell et al., 1993; Li and Burgers, 1994).

The picture of a sliding clamp that tracks DNA while holding the polymerase onto the template is consistent with the results of elegant biochemical analyses of the interactions with DNA of β subunit, gene 45 protein, and recently, PCNA (Nossal and Alberts, 1984; Stukenberg et al., 1991, 1994; Jarvis et al., 1991; Burgers and Yoder, 1993; Tinker et al., 1994). Using radioisotope labeling or cross-linking, it has been shown that the clamps do not load onto closed DNA without assistance from the ATPase subunits of the polymerase. Once loaded onto duplex DNA, they can track along it unless prevented by obstructions, such as bound transcription factors or hairpin structures in single-stranded DNA. Stable complexes of these proteins and circular DNA can be dissociated by linearizing the DNA and allowing the protein to fall off the ends.

An insightful analysis of such experiments carried out for the E. coli Pol III system led to the suggestion that the β subunit might form a toroidal structure that could encircle DNA (Stukenberg et al., 1991). This was confirmed by the determination of the crystal structure of the β subunit, which revealed that a β dimer forms a closed circular ring with a hole in the middle that is large enough for the passage of duplex DNA with no steric hindrance (Kong et al., 1992). Several other features of the structure, most strikingly the electrostatic potential generated by the molecule, make it likely that the protein functions as a ring through which DNA is threaded, thus serving as a DNA-bound platform for the polymerase machinery. Evidence for this model is provided by electron microscopic visualization of the DNA complexes formed by the T4 gene 45 protein (the clamp) and the gene 44/62 proteins (the clamp loaders), which reveals strands of DNA encircled by disk-shaped structures with dimensions matching those of the β subunit dimer (Gogol et al., 1992).

There is no significant sequence similarity between PCNA (258 residues), gene 45 protein (227 residues), and the β subunit (366 residues) and, therefore, no reliable conclusions could be drawn regarding possible structural relationships between them. However, an unexpected feature of the three-dimensional structure of the β subunit immediately suggested such a connection. Despite the lack of internal symmetry in the sequence, each molecule of β subunit consists of three domains of identical topology that interact to form an approximately 6-fold symmetric dimer (Kong et al., 1992). This structural modularity and the fact that PCNA and T4 gene 45 protein are approximately two-thirds the size of β subunit suggested that the former proteins might form trimeric rings, with each monomer of the protein contributing two domains that are similar to those found in the β subunit. The sequences of PCNA and gene 45 protein were aligned with the sequences of the domains in the β subunit so as to preserve the hydrophobic nature of the internal core and to maintain net positive charge surrounding the hole in the ring (Kong et al., 1992). However, the extremely low level of sequence identity (~10%) between the structures made the sequence alignment and the corresponding structure prediction rather unreliable.

We have now determined the crystal structure of PCNA from the yeast Saccharomyces cerevisiae by multiwavelength anomalous diffraction (MAD), and we show that it does indeed form a trimeric ring with close topological symmetry to the structure of the β subunit dimer, as predicted previously. The accuracy of the earlier prediction appears, however, to have been more a triumph of intuition than of reason since we find that the sequence alignment

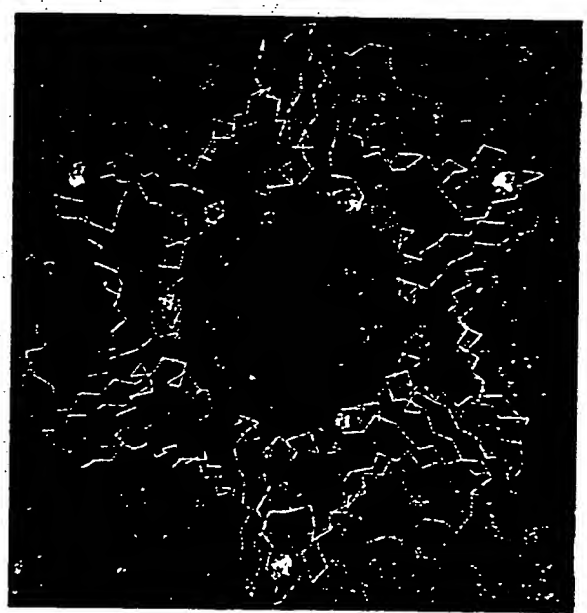


Figure 1. Electron Density Map Calculated Using Phases Derived from MAD Measurements

The red lines represent electron density contours at the 2.2 σ level in an electron density map calculated using amplitudes and phases derived solely from the MAD experiment and modified using SQUASH (Zhang and Main, 1990). This map, calculated at 3 Å resolution, incorporates no prior information about the atomic structure and shows clearly that a PCNA trimer forms a closed ring with a hexagonal outline. For reference, the Cα backbone of the final refined model is shown in yellow. The crystallographic 3-fold axis passes through the center of the ring. Extra electron density features at the upper left and upper right and to the lower right are due to three symmetry-related rings that are packed at an angle to the central one. The mercury atoms (two per monomer) are shown as blue spheres. This figure was generated by program "O" (Jones et al., 1991).

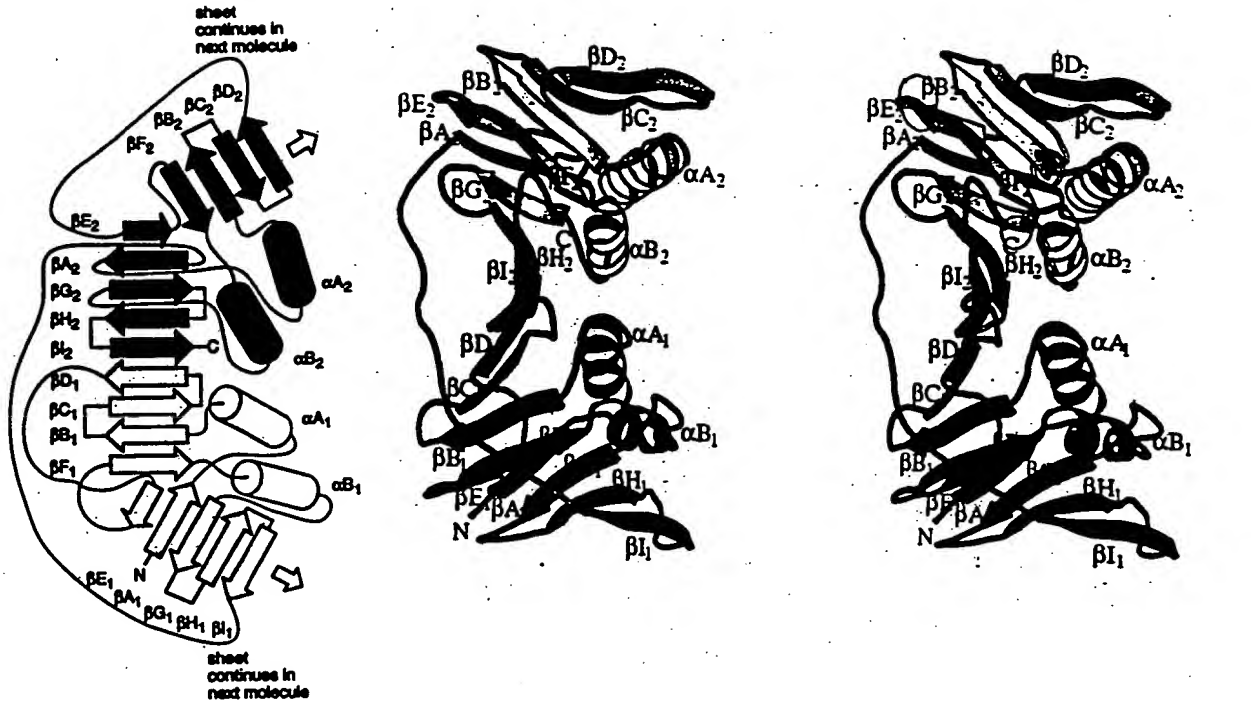


Figure 2. Schematic Diagram and Stereodiagram of a Monomer of PCNA

(Left) Schematic diagram of secondary structural elements in a monomer of PCNA. α Helices are shown as cylinders, and strands of β sheet are shown as arrows. Elements within the two topologically identical domains are differentiated by shading. The secondary structural elements in the first domain are sequentially labeled βA_1, βB_1, etc., and the elements in the second domain are labeled βA_2, βB_2, etc. Loops connecting elements of secondary structure are identified uniquely by the labels of the strands or helices. For example, the A_1B_1 loop is the connection between αA_1 and βB_1. The lengths of the connecting loops are greatly distorted owing to the inability to represent the curvature of the β sheets. The residue numbers of residues at the boundaries of the secondary structural elements are shown in Figure 6. This schematic diagram is oriented similarly to that in the right panel.

(Right) Stereodiagram of the three-dimensional structure of one monomer. The N-terminal domain is shown in green, and the C-terminal domain is shown in yellow. The interdomain connecting loop is shown in red. This stereodiagram was generated using MOLSCRIPT (Kraulis, 1991).

on which the prediction was based (Kong et al., 1992) is largely incorrect. The crystallographic analysis presented here now provides firm evidence that the close correspondence between form and function that was observed in the prokaryotic DNA polymerase processivity factor is preserved in the corresponding eukaryotic protein, and it provides an accurate structural model for the detailed investigation by mutagenesis of the various interactions mediated by PCNA.

Results and Discussion

Structure Determination

The crystal structure of PCNA from S. cerevisiae was determined using phases derived from MAD (Hendrickson, 1991), measured for a mercury complex of the protein. The experimentally derived phases resulted in an electron density map of high quality (Figure 1), on the basis of which a model for the PCNA structure was built and refined against data to 2.3 Å resolution to a final R value of 18.8%. Mercuration of PCNA results in crystals that diffract significantly better than those for the unmercurated protein (Krishna et al., 1994), and the structure of the mercury complex will be used for all the analysis that follows. The structure of unmercurated PCNA has been refined at 3 Å resolution, and it confirms that no significant conformational changes are induced as a result of mercuration (see Experimental Procedures).

Structure of the Trimeric Ring and the Component Domains

The crystal structure of PCNA reveals a closed circular ring, which results from tight association between three molecules that are related by a crystallographic 3-fold axis (Figure 1). Internal symmetry in each monomer of PCNA leads to approximate hexagonal symmetry in the trimer (Figures 1, 2, and 3). This internal symmetry is not reflected in the sequence of PCNA, but is obvious in the secondary structural elements and in the fold of the polypeptide chain (see below). The ring formed by PCNA is strikingly similar to that formed by the β subunit in terms of shape, size, and internal architecture. In both structures, an outer layer of 6 β sheets forms a circular collar that supports 12 α helices lining the inner surface of the ring (Figure 3A). The notable difference between PCNA and β subunit is that the former is a trimer (258 residues, and 28.916 kDa per monomer), whereas the latter is a dimer (366 residues, and 40.583 kDa per monomer). Hydrodynamic (Bauer and Burgers, 1988) and dynamic light scattering experiments (data not shown) indicate that, in

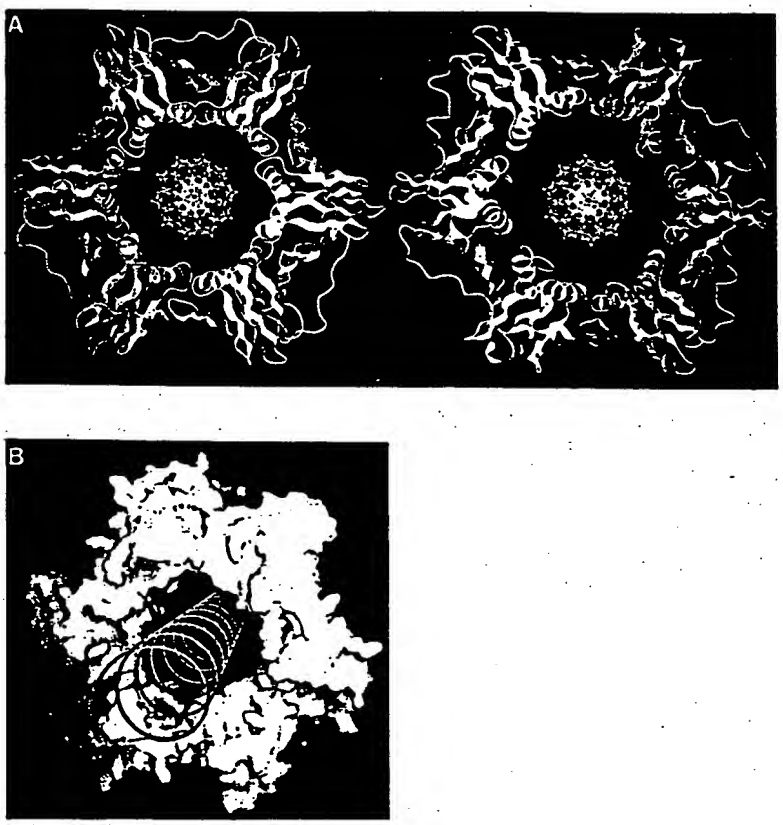


Figure 3. Computer-Generated Images of PCNA Trimer and β Subunit Dimer

(A) Ribbon representations of the polypeptide backbones of PCNA and β subunit, with hypothetical duplex DNA. In this schematic representation of a trimer of PCNA (left) and a dimer of β subunit (right), strands of β sheet are shown as flat ribbons, and α helices are shown as spirals. Individual monomers within each ring are distinguished by different colors. A hypothetical model of duplex B-form DNA is placed in the geometrical center of each structure to illustrate the hypothesis that the rings encircle duplex DNA. This figure was generated using QUANTA (Molecular Simulations). (B) Molecular surface of the PCNA trimer. The solvent-accessible surface of the trimer is displayed, with the surface colored differently for each of the monomers. The phosphate backbone of a strand of standard B-form DNA is indicated by the red spiral. Only 1 of the 2 strands of duplex DNA is shown for clarity. This figure was generated using GRASP (Nicholls et al., 1991).

solution, the molecular mass of PCNA is 80 kDa–85 kDa, corresponding to a trimer and suggesting that the crystal structure reflects the state of the molecule in solution.

The close structural correspondence between the rings formed by trimers of PCNA and dimers of β subunit arises because there are six topologically identical domains in each of the rings. Each PCNA monomer contributes two domains to the trimeric ring, as predicted previously, whereas each β subunit monomer contributes three to the dimeric ring (Kong et al., 1992). The domains consist of two antiparallel β sheets that approach each other closely at one edge, as in a β sandwich, but that are splayed apart at the other by two α helices that pack against the core of hydrophobic residues between the sheets (see right panel, Figure 2). The resulting angle between the sheets is about 45°, and the structure resembles a wedge with the two α helices fronting the blunt edge of the wedge.

The two domains in PCNA are joined firmly together by the formation of an extended β sheet that forms a contiguous surface across the interdomain boundary. The other β sheets in the domain are extended across the intermolecular boundaries. The curvature of the sheets and the angle between them results in circular symmetry: a rotation of 60° about an axis passing through the center of the ring rotates the N-terminal domain of a monomer into close correspondence with the C-terminal domain (rms deviation of 1.1 Å for 55 Cα atoms within secondary structural elements). A rotation of 120° maps the N-terminal domain of one monomer to the corresponding domain in the next monomer in the trimeric ring, a consequence of the crystallographic symmetry. This results in a 9-stranded antiparallel β sheet being formed across the intermolecular interface, with interactions between the C-terminal domain of one monomer and the N-terminal domain of the next one being topologically similar to those between the N- and C-terminal domains of the same monomer.

The modular construction of PCNA that is apparent in the repetitive arrangement of the domains also manifests itself at the level of the secondary structural elements, which form a 2-fold symmetric pattern within each domain. As in the β subunit, the entire ring results from the 12-fold repeat of a β–α–β–β–β motif around a circle (for example, βA–αA–βB–βC–βD or βF–αB–βG–βH–βI in either domain). A short strand (βE in both domains of PCNA) is absent in the β subunit and can be considered an insertion in the loop connecting βD to βF. The structural correspondence between the two halves of a domain is only approximate: rotation by 180° about an axis perpendicular to the 3-fold axis of the ring and passing between the two helices results in a rough alignment of the two halves, with positional deviations of as much as 3 Å–4 Å for Cα atoms in corresponding residues.

The connecting loop between the N- and C-terminal domains is particularly interesting as it might be important for the maintenance of the relative orientation between the two domains in a monomer. The last strand of the N-terminal domain (βI_1) is the farthest from the first strand (βA_2) of the C-terminal domain, necessitating a long con-

Figure 4. Electrostatic Potential Maps for a PCNA Trimer

These maps were generated and displayed using the program GRASP. The side chains of lysine, arginine, asparagine, and glutamate residues were assigned single positive or negative charges as appropriate, and all other residues were considered neutral. The electrostatic potential was calculated using a uniform dielectric of 80 and an ionic strength corresponding to 0.1 M NaCl for the solvent and a dielectric of 2.0 for the protein interior. Orthogonal two-dimensional slices of the electrostatic potential map are shown, passing through the center of the ring. The brightest red and blue colorations correspond to electrostatic potentials of −2.0 and +2.0 k_bT/e, respectively, where k_b is the Boltzmann constant, T is the temperature, and e is the charge of the electron. The polypeptide backbone of PCNA is indicated by an open tube.

necting loop that runs across strands that make up the interdomain β sheet (this loop is marked in red in right panel, Figure 2). Seven hydrophobic residues in this loop extend toward the surface of the underlying β sheet and interact with hydrophobic side chains from both domains. In addition, polar and charged groups on the loop interact either directly or via water molecules with residues in the β sheet (resulting in five direct hydrogen bonds, seven water-mediated hydrogen bonds, and one ion pair). The nature of this cross-back connection between domains is conserved in the β subunit (which has one additional connecting loop owing to the presence of three domains within each monomer) and evokes comparison with the bands that serve to keep the staves of a barrel in place.

Intermolecular Interactions in the PCNA Trimer

The central element of the intermolecular interface is an antiparallel interaction between strands βD_2 in one molecule and βI_1 in the next, which results in a total of eight main chain amide-to-carbonyl hydrogen bonds across the intermolecular interface. This interaction is analogous to that formed by strands βD_1 and βI_2 at the interdomain interface within a monomer, with the same number of bridging hydrogen bonds. The importance of these interactions at the trimer interface is illustrated by a yeast PCNA mutant (Ser-115→Pro in βI_1) that is cold sensitive for growth. The introduction of a proline in the βI_1 strand is likely to disrupt the hydrogen bonding interactions at the intermolecular interface and, indeed, the monomer–trimer equilibrium of the mutant PCNA is strongly shifted toward the monomer form (Burgers et al., unpublished data).

Two helices, αA_2 in one molecule and αB_1 in the next, pack against each other at the interface. The hydrophobic core of each domain is formed by the packing of these (and other) helices against the β sheets, and the formation of the intermolecular interface leads to the burial of one edge of the hydrophobic core in each domain. Specifically, the side chains of Ile-78, Ala-112, Tyr-114, and Leu-116 in one molecule and Leu-154, Val-180, Ile-181, and Ile-182 in the other are solvent exposed in separated monomers, but are buried in the trimer. In addition, the side chains of Ile-175 and Phe-185 undergo partial burial upon formation of the interface. The total surface area buried at each interface of the PCNA trimer is 1500 Å², which is comparable to the 1400 Å² of surface buried at each interface in the β subunit dimer (calculated using a 1.4 Å radius probe).

One characteristic feature of the β subunit dimer is the presence of six potential ion pairs at each interface, two of which are buried, with all the positively charged groups contributed by one molecule and all the negative groups by the other. This polarization of the ends of the monomer would ensure formation of a head-to-tail dimer, with resulting asymmetry in the faces of the ring (Kong et al., 1992). Head-to-tail interactions also occur in the PCNA trimer, although ion pairing is not a prominent feature of the PCNA interface. Only one ion pair is observed at the interface, between Asp-150 in αA_2 and Arg-110 in βI_1. The intermolecular hydrogen bonding network is less extensive in the β subunit, which has only four hydrogen bonds between strands βD_2 and βI_1, compared with eight in PCNA. Despite these differences, the intermolecular interactions in both PCNA and β subunit are substantial, suggesting that energy would have to be expended to break the ring.

Potential Interactions with DNA

The structure of PCNA, like that of the β subunit, has been determined in the absence of DNA. Indeed, the ability of PCNA to slide along duplex DNA (Tinker et al., 1994; M. O'Donnell, personal communication) makes it unlikely that a protein–oligonucleotide complex that is sufficiently stable for crystallography could be generated. Despite the inability to visualize a PCNA–DNA complex crystallographically, the toroidal shape of PCNA suggests immediately that it can fulfill its DNA-tracking function by encircling the duplex (Figure 3). Further examination of the electrostatic potential generated by the molecule and the size of the hole in the middle of the ring provides more evidence to support this hypothesis.

Like β subunit and gene 45 protein, PCNA is highly

acidic. A PCNA trimer contains 81 aspartate, 57 glutamate, 54 lysine, 24 arginine, and 6 histidine residues, which results in a net charge of −60 if the histidines are assumed to be neutral. This is even more negative than the β subunit ring, which carries a net charge of −22. However, like the β subunit, the distribution of charge on the PCNA ring results in a region of positive potential surrounding the inner surface of the ring and the hole in the middle. The qualitative effect of the distribution of charged residues can be seen by calculating a map of the electrostatic potential surrounding the molecule (Sharp and Honig, 1990; Nicholls et al., 1991). Visualization of the map (Figure 4) shows that PCNA generates a large region of negative electrostatic potential around it, consistent with the high negative charge on the ring. The exception is the central channel enclosed by the 12 α helices. This is a region of positive potential, indicating that the phosphate backbone of DNA can pass through the ring without electrostatic repulsions. The positive potential is a consequence of nine lysine and arginine residues within each monomer, in or adjacent to the helices that encircle the inner surface. Alignment of a variety of PCNA sequences shows that lysine or arginine is conserved at each of these nine positions, indicating a functional importance for the charge distribution.

As in the β subunit, each of the central helices is oriented orthogonal to the local direction of the phosphate backbone of standard B-form double helical DNA placed in the center of the ring. This arrangement of helices would prevent PCNA from interacting closely with the grooves of DNA and may restrict the DNA–protein interactions to nonspecific contacts with the phosphate backbone. In order to obtain a sense of the dimensions of the rings formed by PCNA and β subunit, we have computed the rotationally averaged atomic density. Consider each of the rings to be oriented horizontally, such that the rotational symmetry axis is vertical. Moving horizontally outward from the center of the ring, the density of atoms starts to rise at about 12 Å–18 Å from the center for PCNA and at about 14 Å–20 Å for the β subunit. A precise internal diameter is difficult to define because of the variable lengths and conformations of side chains lining the inner surface. Taking the distance at which the atomic density rises to half maximal as a reference value, we obtain internal diameters of approximately 34 Å for PCNA and 38 Å for β subunit. The cross-sectional diameter of the B-form DNA double helix, estimated the same way, is approximately 18 Å (~21 Å in the case of A-form DNA) and, thus, duplex DNA can pass through the center of the ring formed by either PCNA or β subunit without hindrance (see Figure 3). However, the somewhat smaller hole in PCNA, with some side chain atoms extending to within ~12 Å of the center of the ring, means that direct contact between phosphate oxygens and protein side chains may occur even when the DNA is placed in the geometric center of the ring.

Analysis of the density of atoms also leads to estimates of ~80 Å for the external diameter and ~30 Å for the thickness of the rings formed by both PCNA and β subunit. Cryoelectron microscopic images of DNA and the accessory proteins of the T4 phage DNA polymerase reveal a number of disk shaped objects, referred to as "hash marks", that appear to encircle DNA (Gogol et al., 1992). These hash marks or disks are approximately 85 Å wide and 25 Å–30 Å thick and resemble PCNA or the β subunit in cross-section (see Figure 5a of Gogol et al., 1992). Assuming that the disks represent the gene 45 protein, these electron microscopic images are a direct observation of a processivity factor encircling DNA.

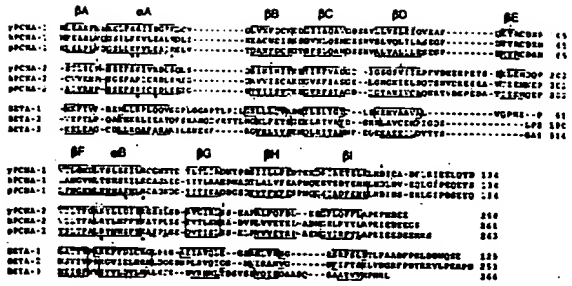


Figure 5. Alignment of PCNA and β Subunit Sequences

The N-terminal domains of S. cerevisiae PCNA, human PCNA, and plant (rice, Oryza sativa) PCNA are labeled yPCNA-1, hPCNA-1, and pPCNA-1, respectively. Other mammalian PCNA sequences are almost identical to that of human PCNA, and the plant sequence was chosen to illustrate sequence variability. The corresponding C-terminal domains are labeled yPCNA-2, hPCNA-2, and pPCNA-2. The three domains of β subunit are labeled BETA-1, BETA-2, and BETA-3. The yPCNA and β subunit sequences are aligned based on the two three-dimensional structures, and the secondary structural elements are boxed and labeled. In domain 2 of PCNA, αB is extended into a 3_{10} helix. This is indicated by broken lines. For yPCNA and the β subunit, the stippled bars indicate residues that are inaccessible to solvent in the crystal structures of the trimer and the dimer, respectively. No consideration of sequence conservation was used in the selection of these residues. However, for the hPCNA and pPCNA sequences, the stippled bars indicated residues that are similar in property to those that are buried in yPCNA. The asterisks under the PCNA sequences indicate the locations of conserved basic residues that contribute to the positive electrostatic potential in the center of the ring. The numbers to the right are the last residue numbers in each row, and the distance between the vertical lines is 10 residues in the yPCNA sequence.

Structure-Based Sequence Comparison between PCNA and β Subunit

An unambiguous alignment of the sequences of PCNA and β subunit was generated for each of the domains on the basis of the known crystal structures (Figure 5). The resulting sequence identity between the two domains of PCNA is ~15%, which is below the level required for the reliable alignment of sequences in the absence of structural information (Sander and Schneider, 1991). Nevertheless, the two domains are closely superimposable, with an rms deviation of 0.9 Å in Cα positions for 55 residues that form the core secondary structural elements (residues that are in corresponding β strands or α helices in each of the domains of PCNA and β subunit). There is no discernible pattern of sequence identity between the two domains. Rather, the changes in the sequence are distributed throughout the structure, with compensatory changes in the hydrophobic core resulting in preservation of the architecture of the domain.

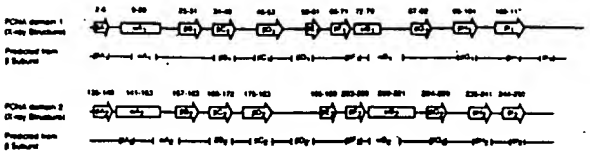


Figure 6. Delineation of the Secondary Structural Elements of PCNA

The identification of residues with β strands and α helices was made on the basis of hydrogen bonding patterns in the crystal structure of PCNA, and the residue numbers of the boundary residues are shown for the two PCNA monomers. Underneath each domain, the previous prediction of the secondary structure (Kong et al., 1992) is indicated. Note that this prediction is out of register with the correct secondary structural boundaries in most cases.

The level of similarity observed between the domains of PCNA is comparable to that observed between PCNA and β subunit or to that seen within the β subunit. The pairwise sequence identities range from 6% (comparing domains 2 and 3 of β subunit) to 15% (the two domains of PCNA). The rms deviations in Cα positions for the core secondary structural elements range from 0.9 Å–1.7 Å. The structure-based alignment of the PCNA and β subunit sequences (Figure 5) shows that each of the β strands and α helices of the β subunit is preserved in PCNA. As in the comparison between the domains of PCNA, there is no pattern of sequence identity between PCNA and β subunit. The hydrophobic nature of the residues in the cores of the domains is indeed preserved, but the corresponding sequence changes are so extensive that simple sequence alignment in the absence of structural information on both proteins would be very difficult. The previous alignment of the sequences of PCNA and β subunit (Kong et al., 1992), based only on inspection of the structure of the latter, is consequently in serious error (Figure 6). Only 5 of the 18 β strands were predicted in the correct positions, and 2 of the 4 α helices are incorrectly aligned. The βE strands of PCNA were not predicted since they are absent in the β subunit. We note that this alignment relied only on manual inspection and intuition and that it would be interesting to know how well computer-based predictions might perform in this case.

The sequence alignment between yeast PCNA and β subunit has been extended to include the sequences of human and a plant PCNA as representative of the diversity of PCNA sequences. In contrast with the alignment of PCNA with β subunit, this extension to other PCNAs is very reliable. Yeast PCNA is 35% and 39% identical in sequence to the human and plant versions, respectively, which is sufficient to ensure that the secondary structural framework will be closely preserved (Sander and Schneider, 1991). There are hardly any insertions or deletions between these sequences, which share clearly recognizable patterns of sequence identity throughout their lengths. Consequently, the structure of yeast PCNA presented here is a reliable guide to the three-dimensional chain fold of human PCNA, in which there is considerable interest.

What of the structure of the T4 phage gene 45 protein? Given the close correspondence between the structures of PCNA and β subunit, the electron microscopic images of the T4 accessory proteins (Gogol et al., 1992), and the fact that gene 45 protein is a trimer (Jarvis et al., 1989), it is very likely that gene 45 protein will adopt a similar architecture. Nevertheless, given the difficulties encountered in generating an accurate sequence alignment between PCNA and β subunit, it would be best to postpone further discussion until the three-dimensional structure of gene 45 protein is determined experimentally.

Interaction of PCNA with Other Molecules

During phases of DNA replication or repair, PCNA interacts with the RFC complex, with Pol δ, and perhaps with other components of the replication or repair machineries. At other times; PCNA is found complexed to the cell cycle proteins p21, cyclin-dependent kinase, and various cyclins (Zhang et al., 1993). Relatively little is known about the specific regions of PCNA that are important for mediating these interactions. Deletion mutagenesis has been used to map the regions of PCNA that potentially interact with antibodies (Brand et al., 1994) and with D-type cyclins (Matsuoka et al., 1994), but the relatively large deletions used makes it difficult to interpret these results in terms of localized regions in the folded structure of PCNA. The two most prominent loops in PCNA, the interdomain connector (residues 118–135; see right panel, Figure 2 and Figure 3A) and the D_2E_2 loop in the second domain (residues 183–195), resemble protuberant "handles" on the trimer and are likely sites for intermolecular interaction. Not surprisingly, these loops are highly antigenic (Roos et al., 1993; Brand et al., 1994).

Five mutations in yeast PCNA have been identified that suppress the phenotype of mutations in the large subunit of the RFC complex (the cdc44 gene product) (McAlear et al., 1994). Based on the previous alignment of PCNA and the β subunit (shown here to be incorrect), it appeared that the five mutations clustered in one region of the predicted structure of PCNA. We have reexamined the locations of these mutations in the crystal structure of PCNA and find that all occur at positions that are completely or partially buried in the structure of the trimer. The mutations are therefore likely to cause changes in the structure of PCNA, particularly in the vicinity of the β sheet that extends across the intermolecular interface. Two of the mutations are in strands that form this sheet (Ala-112→Thr in βI_1 and Val-203→Ala in βF_2). Two other mutations (Leu-151→Ser in αA_2 and Ser-135→Phe in βA_2) involve side chains that pack against this sheet. The fifth mutation (Gly-69→Asp in βF_1) is particularly interesting since the Cα atom of Gly-69 is tightly packed between the Cα–Cβ bonds of two side chains (Met-119 and Ile-121) that are part of the interdomain crossover loop. This loop appears to play an important structural role in maintaining the integrity of the ring, and mutations at position 69 are likely to cause significant changes in the structure of the loop in the vicinity of the intermolecular β sheet. That these mutations rescue phenotypes associated with defective RFC complexes suggests that the large subunit of RFC might interact with PCNA in the vicinity of the trimer interface or, perhaps more likely, that the mutations might affect the monomer–

Table 1. X-Ray Data Collection Statistics

Statistical Parameters	Mercurated PCNA Reference Data set	Mercurated PCNA MAD Phasing			Unmodified PCNA
		λ_1 = 1.0095 Å	λ_2 = 1.0063 Å	λ_3 = 0.9920 Å	
Resolution (Å)	30.0–2.3	30.0–2.5	30.0–2.5	30.0–2.5	30.0–3.0
Unit cell (Å)	121.2	121.4	121.4	121.4	121.7
Number of observed reflections	300,303	120,602	121,809	126,513	95,433
Number of unique reflections	26,608	20,934	20,935	20,932	12,273
Completeness of all data (%)	93.7	73.6	72.8	74.3	86.3
Overall $I/\sigma(I)$	24.3	16.5	16.6	16.5	18.1
Completeness of outer shell (Å)	92.7(2.34–2.30)	56.8%(2.54–2.50)	56.8%(2.54–2.50)	58.5%(2.54–2.50)	80.9%(3.05–3.0)
$I/\sigma(I)$ of outer shell (Å)	5.2(2.34–2.30)	3.2(2.54–2.50)	3.2(2.54–2.50)	3.2(2.54–2.50)	3.9(3.05–3.0)
$R_{symmetry}$ (%)	8.9	6.0	6.6	6.6	6.5

Five different data sets are shown: four for mercurated PCNA and one for unmercurated PCNA. The first column gives statistics for the best data set measured for mercurated PCNA using a rotating anode X-ray source in the laboratory with a copper target. The number of unique reflections is calculated in each case by considering Friedel pairs to be equivalent. The next three columns are for data measured at the synchrotron at the three wavelengths indicated (λ_1, λ_2, λ_3). The last column is for the unmercurated protein, measured in the laboratory. All data collections were carried out using crystals frozen to ~−160°C. $I/\sigma(I)$, average intensity expressed in terms of the standard deviation in the intensity, and $R_{symmetry}$ agreement factor, on intensity, between redundant measurements of the same reflection.

trimer equilibrium of PCNA in a way that compensates for the mutations in the large subunit. The available data do not allow us to distinguish between these alternatives, but the crystal structure should be a valuable resource in designing mutations to test this and other interactions of PCNA.

Conclusion

The crystal structures of PCNA and the β subunit are the only views obtained so far of any of the components of the molecular machines that replicate chromosomal DNA. Unexpectedly, the three-dimensional architectures of these processivity factors are strikingly similar, despite the differences in their oligomerization states and the lack of similarity in their sequences. These structures explain the tight affinity of PCNA and the β subunit for DNA, which is topological, but now bring to the forefront the question of how the associated ATP-dependent auxiliary proteins (γ complex in E. coli and the RFC complex in eukaryotes) act to load these closed rings onto DNA. Sequence homology between the components of these auxiliary proteins (O'Donnell et al., 1993; Li and Burgers, 1994) indicates that some commonality will likely be found in their mode of action in eukaryotes and prokaryotes, consistent with the structural homology in the corresponding processivity factors. Very little is also known at present about the structures of the catalytic subunits of the chromosomal replicases and to what extent their attachment to the processivity factors will resemble each other. The intriguing possibility that PCNA might play a role in coupling the regulation of DNA replication to cell cycle–dependent signals points to another important class of interactions involving PCNA that needs to be understood in structural terms. Continuing analysis of the three-dimensional structures of the components of these replicases is likely to lead to new insights into the basic biological process of DNA replication.

Experimental Procedures

Crystallization and Data Collection

The purification and crystallization of yeast PCNA has been described previously (Bauer and Burgers, 1988; Yoder and Burgers, 1991; Krishna et al., 1994). In brief, large single crystals are obtained from high concentrations (~2 M) of ammonium sulfate at pH 5.6. Crystals grown from solutions containing mercurated PCNA (obtained by reacting PCNA with 0.4 mM $HgCl_2$, followed by dialysis to remove unreacted mercury) diffract significantly better than unmodified PCNA, with measureable X-ray data extending to 2.3 Å and 3.0 Å for the mercurated and unmercurated crystals, respectively (Krishna et al., 1994).

The best X-ray diffraction data for mercurated PCNA were obtained using a single crystal that was flash frozen to −160°C (space group $P2_13$, $a = b = c = 121.2$ Å, one monomer of PCNA in the asymmetric unit) using a Rigaku RAXIS-IIC detector mounted on a rotating anode X-ray source (Table 1). The anomalous difference Patterson map calculated from these data is of high quality, with strong Harker section peaks corresponding to two mercury atoms bound to the protein (peak heights are ~8 σ above the mean, using data to 3 Å resolution). In contrast, isomorphous difference Patterson maps are noisy, with Harker section peaks that are only half as strong as those in the anomalous difference Patterson map, suggesting significant nonisomorphism between crystals of the mercurated and unmercurated protein. The diffraction from crystals of the unmodified protein is substantially weaker than that from mercurated PCNA, and the quality of the data is correspondingly poor (Table 1). We therefore chose not to use data for the unmodified protein in the determination of phases, and we relied instead on anomalous diffraction from the mercury atoms to determine the phases experimentally.

Phase Determination by MAD

A MAD experiment was carried out at the Howard Hughes Medical Institute synchrotron resource (beamline X4A at the National Synchrotron Light Source, Brookhaven National Laboratory [Upton, New York]) (Hendrickson, 1991). One crystal of mercurated PCNA, frozen at −160°C, was used to collect data at X-ray wavelengths corresponding to the inflection point (1.0095 Å), peak (1.0063 Å), and a remote position (0.9920 Å) of the X-ray absorption spectrum of the crystal. The data were measured by the oscillation method, using Fuji imaging phosphor plates. Oscillations of 1.5° were used, with an overlap of 0.3° between successive oscillations. The data collection spanned a total oscillation range of 72°, corresponding to two 36° sectors that are related by a 180° rotation about the oscillation axis. Data for each of the three wavelengths were measured sequentially for a particular oscillation range (and that related by a rotation of 180°) before advancing to the

Table 2. Crystallographic Data

	Observed Ratio of Anomalous Diffraction Differences			Scattering Factors	
Wavelength (Å)	λ_1	λ_2	λ_3	f'	f''
λ_1 = 1.0095	0.070	0.037	0.042	−24.65	7.26
λ_2 = 1.0063		0.078	0.038	−16.12	11.60
λ_3 = 0.9920			0.076	−10.87	9.88

Bijvoet differences from data measured at the three wavelengths indicated are shown in the diagonal elements of the table. Dispersive differences are shown as off-diagonal elements. The values of the scattering factors f' and f'' are listed for each wavelength; f' and f'' represent the values of the real and imaginary components of the anomalous scattering factor. The values at λ_1 and λ_2 were refined using MADLSQ, with the values at λ_3 fixed at their calculated values (W. A. Hendrickson, personal communication).

next one. The resulting images were processed using the program DENZO, and the integrated intensities were scaled and merged using SCALEPACK (Z. Otwinowski, personal communication). The anomalous differences were also merged and reduced to a unique set of indices prior to all further analysis. This increased the completion of the anomalous difference measurements for the unique set of indices while taking advantage of the high symmetry of the cubic space group. The resulting wavelength-dependent statistics for the anomalous differences are shown in Table 2.

The algebraic formulation of Hendrickson was applied to the reduced multiwavelength data to obtain phase probabilities. The mean figures of merit, reported by the program MADABCD (Pahler et al., 1990), are 0.71 (to 3 Å) and 0.64 (to 2.5 Å). The phases were improved by solvent flattening and histogram matching using the program SQUASH (Zhang and Main, 1990), initially at 3 Å resolution and ultimately at 2.5 Å resolution (see below).

The map calculated using the phases obtained from SQUASH revealed unambiguously that PCNA forms a ring-shaped, 3-fold symmetric structure with close topological similarity to the E. coli Pol III β subunit (see Figure 1). The map had clearly resolved side chain features that could be interpreted readily in terms of the sequence of yeast PCNA (Bauer and Burgers, 1990). This revealed that although the topology of the chain fold resembled that of the β subunit, the specific assignment of secondary structural elements in PCNA that had been suggested previously (Kong et al., 1992) was largely incorrect (Figure 6). The sequence of PCNA was therefore built into the electron density map independently (there is one monomer of PCNA in the asymmetric unit of the crystal). An atomic model for approximately 60% of the sequence was built into the first electron density maps using the program "O" (Jones et al., 1991), after which phases calculated using the atomic model were combined with the MAD phases to 2.5 Å using SIGMAA (Read, 1986). A model corresponding to the complete PCNA sequence was built and refined by least-squares optimization and simulated annealing using X-PLOR (Brünger, 1988; Brünger et al., 1990), with the stereochemical parameters of Engh and Huber (1991). The model contains 2147 nonhydrogen atoms, including 117 water molecules. The R value of the final model is 18.8% for 22,454 reflections between 5.0 Å and 2.3 Å with $|F| > 2\sigma(|F|)$. The rms deviation from ideality in bond lengths and angles is 0.012 Å and 1.9°, respectively.

A problem with the phases obtained by the MAD analysis is that they are only 65% complete to 2.5 Å resolution (70% to 3.0 Å), leading to breaks in electron density features. The data set measured in the laboratory is 93% complete to 2.5 Å resolution, with a mean fractional difference of 12.8% to 2.5 Å on |F| with respect to the data measured at the synchrotron. This led us to use the relatively complete set of laboratory structure factors as input to the SQUASH program, with phase probabilities taken from the MAD experiment where available. Unphased reflections were also included in the SQUASH calculation, but with initial phase probability coefficients set to zero. Solvent flattening and histogram matching were used to improve and extend the phases, starting with data to 3.5 Å resolution and proceeding to 2.5 Å in 20 cycles. The resulting phase set is 93% complete to 2.5 Å, and the resulting electron density map is of exceptionally high quality and could be interpreted unambiguously in most regions of the structure, with the exception of certain surface loops.

The improvements in the phase accuracy can be illustrated by comparing the average value of the phase discrepancy between the final phase set calculated from the refined model and the experimental phases at various stages. The average discrepancy is 65° (for data corresponding to Bragg spacings between 5.0 Å–2.5 Å) for the phases resulting from the MAD analysis. Application of SQUASH to these data reduces the discrepancy to 49°. A further improvement is afforded by switching to the complete laboratory data set: the phase discrepancy after SQUASH drops to 41° for those reflections that were phased by the MAD analysis and is 61° for reflections for which no experimental phases were available.

Comparison of the Structures of Mercurated and Unmercurated PCNA

The two mercury atoms that are bound to PCNA are coordinated by 3 of the 4 cysteine residues in the protein. The first mercury cross-links the side chains of Cys-30 and Cys-62 by forming covalent bonds with the sulfur atoms of both residues. The second mercury is bound to Cys-22. The occupancies of the mercury sites were estimated by computing the R value as a function of occupancy and were set to 1.0 and 0.6 for the two sites, respectively. The sulfur–mercury distances were restrained to 2.55 Å during refinement, the value obtained by averaging sulfur–mercury bond lengths observed in several small molecule structures (Wells, 1984). The first mercury cross-links cysteine residues that are on opposite sides of the β sandwich of domain 1. This internally cross-linked region is close to the 3-fold rotation axis of the crystal and is packed close to two symmetry-related copies of the same region in different PCNA trimers. Stabilization of the structure in this region may contribute toward improving the crystalline order. The second mercury atom is bound near the inner surface of the ring and is quite distant from crystal contacts.

The final refined model for mercurated PCNA was used to initiate least-squares refinement for a model for the unmercurated protein at 3.0 Å resolution, which was followed by simulated annealing and manual adjustment of the model. The present R value is 22.4%, with 10 water molecules included. The rms deviation from ideality in bond lengths and angles is 0.013 Å and 2.1°, respectively. Comparison of the structures of mercurated and unmercurated PCNA shows that no gross changes in the structure occur upon mercuration. The rms deviation in Cα positions, after superimposing the monomers, is 0.66 Å overall. The higher resolution of the X-ray data for mercurated PCNA makes it the more accurately determined structure, and it is the model used in all the analysis presented in this paper.

Acknowledgments

We are grateful to Mike O'Donnell for initiating this collaboration, Jacqueline Gulbis and Xiaodong Wu for help with synchrotron data collection, and Craig Ogata and Wayne A. Hendrickson for providing access to and support at beamline X4A. We thank Stephen Burley, Jonathan Goldberg, Zvi Kelman, Joseph P. Kim, Ismail Moarefi, Mike O'Donnell, and Bill Weis for helpful suggestions. This work was supported in part by grants from the National Institutes of Health to P. B. (GM32431) and to J. K. (GM45547). Beamline X4A at the National Synchrotron Lightsource in Brookhaven, New York is supported by the Howard Hughes Medical Institute. Atomic coordinates are being deposited in the Protein Databank and are also available by e-mail (kuriyan@rockefeller.edu).

Received October 3, 1994; revised November 10, 1994.

References

Bauer, G. A., and Burgers, P. A. (1988). The yeast analog of mammalian cyclin/proliferating-cell nuclear antigen interacts with mammalian DNA polymerase δ. Proc. Natl. Acad. Sci. USA *85*, 7506–7510.

Bauer, G. A., and Burgers, P. M. J. (1990). Molecular cloning, structure and expression of the yeast proliferating cell nuclear antigen gene. Nucl. Acids Res. *18*, 261–265.

Brand, S. R., Bernstein, R. M., and Mathews, M. B. (1994). Autoreactive epitope profiles of the proliferating cell nuclear antigen define two classes of antibodies. J. Immunol. *152*, 4120–4128.

Bravo, R., Frank, R., Blundell, P. A., and Macdonald-Bravo, H. (1987). Cyclin/PCNA is the auxiliary protein of DNA polymerase delta. Nature *326*, 515–517.

Brünger, A. T. (1992). X-PLOR, Version 3.1 Manual (New Haven, Connecticut: Yale University Press).

Brünger, A. T., Krukowski, A., and Erickson, J. W. (1990). Slow-cooling protocols for crystallographic refinement by simulated annealing. Acta Crystallogr. *A46*, 585–593.

Burgers, P. M. J. (1991). *Saccharomyces cerevisiae* replication factor C. II. Formation and activity of complexes with proliferating cell nuclear antigen and with DNA polymerases δ and ε. J. Biol. Chem. *266*, 22698–22706.

Burgers, P. M. J., and Yoder, B. L. (1993). ATP-independent loading of the proliferating cell nuclear antigen requires DNA ends. J. Biol. Chem. *268*, 19923–19926.

Cha, T.-A., and Alberts, B. M. (1988). *In vitro* studies of the T4 bacteriophage DNA replication system. In Eukaryotic DNA Replication, T. Kelly and B. Stillman, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press), pp. 1–10.

Engh, R. A., and Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. Acta Crystallogr. *A47*, 392–400.

Flores-Rozas, H., Kelman, Z., Dean, F. B., Pan, Z.-Q., Harper, J. W., Elledge, S. J., O'Donnell, M., and Hurwitz, J. (1994). Cdk-interacting protein 1 directly binds with proliferating cell nuclear antigen and inhibits DNA replication catalyzed by the DNA polymerase δ holoenzyme. Proc. Natl. Acad. Sci. USA *91*, 8655–8659.

Gogol, E. P., Young, M. C., Kubasek, W. L., Jarvis, T. C., and von Hippel, P. H. (1992). Cryoelectron microscopic visualization of functional subassemblies of the bacteriophage T4 DNA replication complex. J. Mol. Biol. *224*, 395–412.

Hendrickson, W. A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron data. Science *254*, 51–58.

Jarvis, T. C., Paul, L. S., and von Hippel, P. H. (1989). Structural and enzymatic studies of the T4 DNA replication system. I. Physical characterization of the polymerase accessory protein complex. J. Biol. Chem. *264*, 12709–12716.

Jarvis, T. C., Newport, J. W., and von Hippel, P. H. (1991). The processivity of the DNA polymerase of bacteriophage T4 by the polymerase accessory proteins. J. Biol. Chem. *266*, 1830–1840.

Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. *A47*, 110–119.

Kong, X.-P., Onrust, R., O'Donnell, M., and Kuriyan, J. (1992). Three-dimensional structure of the β subunit of E. coli DNA polymerase III holoenzyme: a sliding DNA clamp. Cell *69*, 425–437.

Kornberg, A., and Baker, T. A. (1991). DNA Replication, Second Edition (New York: W. H. Freeman).

Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. *24*, 946–950.

Krishna, T. S. R., Kong, X.-P., Gary, S., Burgers, P., and Kuriyan, J. (1994). Crystallization of proliferating cell nuclear antigen (PCNA) from *Saccharomyces cerevisiae*. J. Mol. Biol. *241*, 265–268.

Kuriyan, J., and O'Donnell, M. (1993). Sliding clamps of DNA polymerases. J. Mol. Biol. *234*, 915–925.

Lee, S.-H., Kwong, A. D., Pan, Z.-H., and Hurwitz, J. (1991a). Studies on the activator 1 protein complex, an accessory factor for proliferating cell nuclear antigen-dependent DNA polymerase δ. J. Biol. Chem. *266*, 594–602.

Lee, S.-H., Pan, Z.-Q., Kwong, A. D., Burgers, P. M. J., and Hurwitz, J. (1991b). Synthesis of DNA by DNA Polymerase ε *in vitro*. J. Biol. Chem. *266*, 22707–22717.

Li, X., and Burgers, P. M. (1994). Molecular cloning and expression of the *Saccharomyces cerevisiae RFC3* gene, an essential component of replication factor C. Proc. Natl. Acad. Sci. USA *91*, 868–872.

Matsuoka, S., Yamaguchi, M., and Matsukage, A. (1994). D-type cyclin-binding regions of proliferating cell nuclear antigen. J. Biol. Chem. *269*, 11030–11036.

McAlear, M. A., Howell, E. A., Espenshade, K. K., and Holm, C. (1994). Proliferating cell nuclear antigen (pol30) mutations suppress cdc44 mutations and identify potential regions of interaction between the two encoded proteins. Mol. Cell. Biol. *14*, 4390–4397.

Miyachi, K., Fritzler, M. J., and Tan, E. M. (1978). Autoantibody to a nuclear antigen in proliferating cells. J. Immunol. *121*, 2228–2234.

Morrison, A., Araki, H., Clark, A. B., Hamatake, R. K., and Sugino, A. (1990). A third essential DNA polymerase in S. cerevisiae. Cell *62*, 1143–1151.

Muller, F., Seo, Y. S., and Hurwitz, J. (1994). Replication of bovine papillomavirus type I origin-containing DNA in crude extracts and with purified proteins. J. Biol. Chem. *269*, 17086–17094.

Nicholls, A., Sharp, K. A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins: Struct. Funct. Genet. *11*, 281–296.

Nossal, N. G., and Alberts, B. M. (1984). Mechanism of DNA replication catalyzed by purified T4 replication proteins. In Bacteriophage T4, C. Mathews, ed., (Washington, D. C.: American Society for Microbiology), pp. 71–81.

O'Donnell, M., Onrust, R., Dean, F. B., Chen, M., and Hurwitz, J. (1993). Homology in accessory proteins of replicative polymerases-*E. coli* to humans. Nucl. Acids Res. *21*, 1–3.

Pahler, A., Smith, J. L., and Hendrickson, W. A. (1990). A probability representation for phase information from multiwavelength anomalous dispersion. Acta Crystallogr. *A46*, 537–540.

Podust, V. N., and Hübscher, U. (1993). Lagging strand DNA synthesis by calf thymus DNA polymerase α, β, δ and ε in the presence of auxiliary proteins. J. Biol. Chem. *25*, 841–846.

Prelich, G., Kostura, M., Marshak, D. R., Mathews, M. B., and Stillman, B. (1987a). The cell-cycle regulated proliferating cell nuclear antigen is required for SV40 replication *in vitro*. Nature *326*, 471.

Prelich, G., Tan, C.-K., Kostura, M., Mathews, M. B., So, A. G., Downey, K. M., and Stillman, B. (1987b). Functional identity of proliferating cell nuclear antigen and a DNA polymerase auxiliary protein. Nature *326*, 517–520.

Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. Acta Crystallogr. *A42*, 140–149.

Roos, G., Landberg, G., Huff, J. P., Houghten, R., Takasaki, Y., and Tan, E. M. (1993). Analysis of the epitopes of proliferating cell nuclear antigen recognized by monoclonal antibodies. Lab. Invest. *68*, 204–210.

Sander, C., and Schneider, R. (1991). Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins *9*, 56–68.

Sharp, K. A., and Honig, B. (1990). Electrostatic interactions in macromolecules: theory and application. Annu. Rev. Biophys. Biophys. Chem. *19*, 301–332.

Shivji, M. K. K., Kenny, M. K., and Wood, R. D. (1992). Proliferating cell nuclear antigen is required for DNA excision repair. Cell *69*, 367–374.

Stillman, B. (1994). Smart machines at the DNA replication fork. Cell *78*, 725–728.

Stukenberg, P. T., Studwell-Vaughan, P. S., and O'Donnell, M. (1991). Mechanism of the sliding β-clamp of DNA polymerase III holoenzyme. J. Biol. Chem. *266*, 11328–11334.

Stukenberg, P. T., Turner, J., and O'Donnell, M. (1994). An explanation for lagging strand replication: polymerase hopping among DNA sliding clamps. Cell *78*, 877–887.

Tan, C. K., Castillo, C., So, A. G., and Downey, A. G. (1986). An auxiliary protein for DNA polymerase δ from fetal calf thymus. J. Biol. Chem. *261*, 12310–12316.

Tinker, R. L., Kassavetis, G. A., and Geiduschek, E. P. (1994). Detecting the ability of viral, bacterial and eukaryotic proteins to track along DNA. EMBO J., in press.

Tsurimoto, T., and Stillman, B. (1990). Functions of replication factor C and proliferating-cell nuclear antigen: functional similarity of DNA polymerase accessory proteins from human cells and bacteriophage T4. Proc. Natl. Acad. Sci. USA *87*, 1023–1027.

Waga, S., and Stillman, B. (1994). Anatomy of a DNA replication fork revealed by reconstitution of SV40 DNA replicaton *in vitro*. Nature *369*, 207–212.

Waga, S., Hannon, G. J., Beach, D., and Stillman, B. (1994). The p21 inhibitor of cyclin-dependent kinases controls DNA replication by interaction with PCNA. Nature *369*, 574–578.

Wells, A. F. (1984). Structural Inorganic Chemistry (Oxford: Clarendon Press).

Wold, M. S., Li, J., Weinberg, D. H., Vershup, D. M., Verkeyer, E., and Kelly, T. (1988). Cellular protein required for SV40 DNA replication *in vitro*. Cancer Cells *6*, 133.

Xiong, Y., Zhang, H., and Beach, D. (1993). D type cyclins associate with multiple protein kinases and the DNA replication and repair factor PCNA. Cell *71*, 505–514.

Yoder, B. L., and Burgers, P. M. J. (1991). *Saccharomyces cerevisiae* replication factor C. I. Purification and characterization of its ATPase activity. J. Biol. Chem. *266*, 22689–22697.

Zeng, X. R., Jiang, Y., Zhang, S. J., Hao, H., and Lee, M. Y. (1994). DNA polymerase delta is involved in the cellular response to UV damage in human cells. J. Biol. Chem. *269*, 13748–13751.

Zhang, H., Xiong, Y., and Beach, D. (1993). Proliferating cell nuclear antigen and p21 are components of multiple cell cycle kinase complexes. Mol. Biol. Cell *4*, 897–906.

Zhang, K. Y. J., and Main, P. (1990). The use of Sayre's equation with solvent flattening and histogram matching for phase extension and refinement of protein structures. Acta Crystallogr. *A46*, 377–381.

© 1995 Oxford University Press

Nucleic Acids Research, 1995, Vol. 23, No. 18 3613–3620

Structural and functional similarities of prokaryotic and eukaryotic DNA polymerase sliding clamps

Zvi Kelman* and Mike O'Donnell[1]

Microbiology Department Hearst Research Foundation and [1]Howard Hughes Medical Institute, Cornell University Medical College, 1300 York Avenue, New York, NY 10021, USA

Received June 9, 1995; Revised and Accepted August 15, 1995

ABSTRACT

The remarkable processivity of cellular replicative DNA polymerases derive their tight grip to DNA from a ring-shaped protein that encircles DNA and tethers the polymerase to the chromosome. The crystal structures of prototypical 'sliding clamps' of prokaryotes (β subunit) and eukaryotes (PCNA) are ring shaped proteins for encircling DNA. Although β is a dimer and PCNA is a trimer, their structures are nearly superimposable. Even though they are not hexamers, the sliding clamps have a pseudo 6-fold symmetry resulting from three globular domains comprising each β monomer and two domains comprising each PCNA monomer. These domains have the same chain fold and are nearly identical in three-dimensions. The amino acid sequences of 11 β and 13 PCNA proteins from different organisms have been aligned and studied to gain further insight into the relation between the structure and function of these sliding clamps. Furthermore, a putative embryonic form of PCNA is the size of β and thus may encircle DNA as a dimer like the prokaryotic clamps.

INTRODUCTION

Chromosomal replicases are multiprotein assemblies that polymerize thousands of nucleotides without dissociating from DNA. In three well characterized systems (*E.coli*, eukaryotes and bacteriophage-T4) this remarkable processivity is achieved by a ring-shaped protein ('sliding clamp') that encircles DNA and anchors the polymerase to the template (reviewed in 1,2). The sliding clamp is assembled on DNA by a multiprotein 'clamp loader' apparatus that recognizes a primer terminus and couples ATP hydrolysis to assemble the clamp around DNA (3). In eukaryotes the clamp loader is the 5-subunit RF-C complex (also known as activator-1) and the sliding clamp is the PCNA protein which confers processivity to DNA polymerase δ (Pol δ). The clamp loader of the prokaryotic DNA polymerase III holoenzyme (Pol III) is the 5-subunit γ-complex and the clamp is the β subunit. In the T4 system the clamp loader is the gene 44/62 protein complex (g44/62p) and the clamp is the product of gene 45.

The first indication of the processivity factor's circular shape came from the observation that the β subunit was tightly fastened to circular DNA but easily dissociated from linear DNA, suggesting that β slides off the ends of DNA (4). This hypothesis was enforced by the observation that the exit of β from linear DNA was blocked by proteins bound near the ends of the DNA. The three dimensional structure of the *E.coli* β subunit showed it to be a dimer in the shape of a ring with a central cavity large enough to accommodate duplex DNA (5) (Fig. 1).

The sliding clamp in the T4 system has been observed on DNA by cryoelectron microscopy (6). The clamps not only appeared to encircle the DNA, but their appearance also indicated that they may slide along the DNA. These clamps are presumably gene 45 protein trimers encircling the DNA in a fashion similar to that of the β dimer. Using a linear template it was shown that gene 45 protein can support processive replication in the absence of the clamp loader (7) and that the T4 sliding clamp consists solely of gene 45 protein (8). Interestingly, the gene 45 protein can assemble on a circular template in the absence of the clamp loader, provided a macromolecular crowding agent is present (9).

Evidence for the topological binding of PCNA to DNA came from the observation that PCNA alone can support replication by Pol δ, in the absence of RF-C complex, only when the DNA is linear and has a double-stranded end (10). This result suggests that PCNA can slide along duplex DNA until it reaches the 3′ terminus where it interacts with the polymerase to initiate processive replication. Supporting evidence for PCNA sliding along duplex DNA comes from photocrosslinking experiments in which PCNA can be crosslinked to DNA following assembly around DNA (11). However, crosslinking of PCNA to DNA is not observed upon linearization of the duplex (11). Furthermore, using tritiated human PCNA it has been demonstrated that the protein can be loaded onto nicked circular DNA, but upon linearization of the DNA it slides off the ends, similar to the *E.coli* β subunit (N. Yao, Z. Kelman and M. O'Donnell, unpublished). Proof that PCNA and β are structurally similar was obtained upon solution of the three dimensional structure of the yeast *Saccharomyces cerevisiae* PCNA which showed it to be a trimer with an overall shape very similar to that of *E.coli* β (12) (Fig. 1).

The sliding clamps (β and PCNA) have a 6-fold appearance, yet they are not hexamers. In β the 6-fold appearance results from three globular domains which comprise each monomer, while in PCNA each monomer contains two domains. In both prokaryotes

*To whom correspondence should be addressed at present address: Department of Molecular Biology and Genetics, The Johns Hopkins Medical School, Baltimore, MD 21205, USA

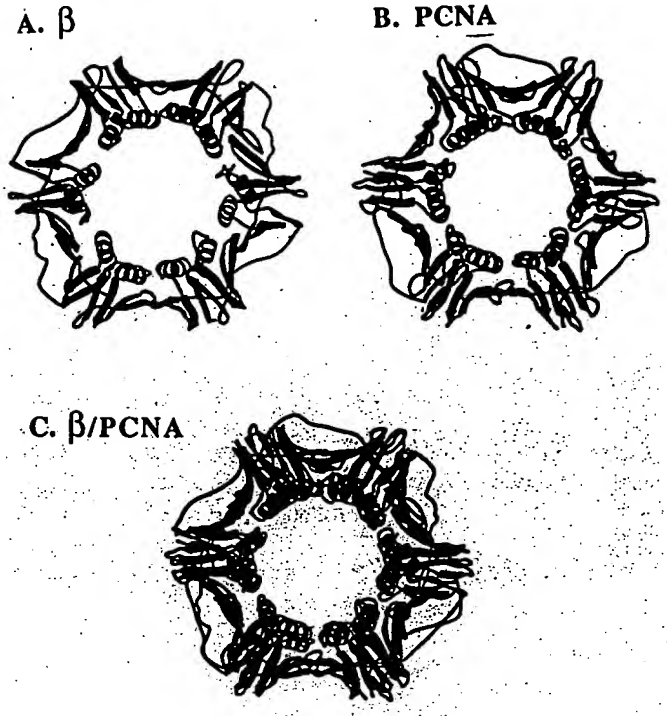


Figure 1. Ribbon representation of the polypeptide backbones of a trimer of PCNA (A) and dimer of β (B). (C) is a superimposition of the PCNA and β rings. Strands of β sheet are shown as flat ribbons and α helices are shown as spirals. The monomeric units within each ring are distinguished by different colors.

and eukaryotes the topological chain fold of each domain is comprised of two α helices supported by eight antiparallel β sheets (nine in PCNA). The domains have the same polypeptide chain fold and are nearly superimposable in three dimensions yet share no significant homology (5,12). The sheet structure is continuous all around the perimeter of the ring and the α helices that line the central cavity are oriented perpendicular to the phosphate backbone of the DNA.

Here we describe the diverse roles played by sliding clamps in nucleic acid metabolism. Also, in an attempt to derive additional insight into the relationship of structure and function in prokaryotic and eukaryotic sliding clamps, the amino acid sequences of several β and PCNA proteins from different organisms have been aligned.

Structural features of the sliding clamp

The three features of the clamp structure that directly relate to its function are: (i) the central cavity where the DNA is located; (ii) the interface between monomers where the protein is opened for loading onto DNA; and (iii) the surface on the ring that interacts with other subunits of the polymerase and with other cellular components. To gain additional insight into the residues important for function, we aligned the amino acid sequences of 11 full or partial prokaryotic β subunits (Fig. 2) and 13 PCNAs (Fig. 3). These sequences were obtained from the databases by searching for known DNA sequences of *dnaN*, the gene that encodes β and genes known to encode for PCNA. To date, only the *E.coli* β and PCNA of yeast and human have been shown functionally to act as DNA sliding clamps.

The β sequences in Figure 2 are split into three parts which correspond to the three globular domains. The sequences in Figure 3 are divided into the two domains of PCNA. Glu and Asp are colored red and Lys and Arg are blue to display more clearly the distribution of charged residues within the protein. The alignments were made such that no gaps were introduced between elements of secondary structure and that even though these proteins have high net negative charges, the residues aligning with the α helices of the central cavity have a net positive charge. These alignments generally follow a third criteria, that positions corresponding to the hydrophobic amino acid residues (Phe, Leu, Ile, Met, Val or Ala) in *E.coli* β and yeast PCNA that are inaccessible to water, are also occupied by hydrophobic residues.

The central cavity

In general these clamps are acidic proteins (Tables 1 and 2). The pI of the PCNAs are slightly lower then the pIs of β subunits and gene 45 protein has an intermediate pI (pI = 4.8). The distribution of charges on the ring is asymmetric. The outer surface has a strong negative electrostatic potential and the inside of the central cavity has a net positive electrostatic potential. Hence, these clamps would be repelled by DNA, but after assembly onto DNA, the ring may even interact favorably with it. A possible function of the negative outside surface may be to increase the specificity of assembly on DNA by preventing the ring from associating with DNA by itself, thus constraining it to enlist the help of the clamp loader and ATP. Alternatively, the negative surface may help destabilize local interaction with DNA. A notable exception to this rule is the β subunit from *B.aphidicola* with a pI of 9.0. Hence, providing this β subunit functions like β of *E.coli*, an overall acidic charge would appear not to be essential to clamp function.

Both β and PCNA have 12 α helices in the central cavity all of which are perpendicular to the DNA. The α helices are long enough to span the major and minor grooves and thus may act as molecular crossbars to prevent the clamp from entering the grooves of DNA during sliding. The α helices have an overall positive potential as a consequence of eight (PCNA) or 12 (β) Lys or Arg residues within the ring. However, there is at least one acidic amino acid (Asp or Glu) in every α helix of β, and in six out of the 12 α helices in PCNA. These negatively charged residues may be needed for a balance to prevent the local interaction between the clamp and DNA. Although the overall structure of the rings are similar, the central cavity in β (35 Å) is larger than the one in PCNA (34 Å) and has a slightly more oval shape (Fig. 1).

The dimer interface

The interfaces between the protomers of β and PCNA are formed by β sheet. The β sheet is part of the continuous layer of sheet structure on the outside of these rings. Three interactions may contribute to stabilization of the interface: (i) hydrogen bonds between the β sheets; (ii) a small hydrophobic core; and (iii) putative ion pairs. In the following sections the amino acids that contribute to the formation of the interface close to the N-terminus will be referred to as the 'head interface' and those near the C-terminus will be referred to as the 'tail interface' (Fig. 4).

Figure 2. Sequence alignment of β subunits from different prokaryotes. The proteins are aligned to show maximum similarity with proteins directly above and below. Identical residues between neighboring proteins have a vertical dash. Gaps were introduced to maximize homology. The sequences are split into three parts which correspond to the three globular domains of the *E.coli* β subunit. The Glu and Asp residues are shown in red and the Lys and Arg residues are blue. To the right are the amino acid numbers at the end of each domain of the *E.coli* β. The secondary structures corresponding to *E.coli* β are indicated above each domain. Amino acids that are conserved in all species are indicated in the bottom row of each domain. The sequences are from the following references: *Buchnera aphidicola* [BUCAP (24)], *Bacillus subtilis* [BACSU (25)], *Micrococcus luteus* [MICLU (26)], *Streptomyces coelicolor* [STRCO (27)], *Pseudomonas putida* [PSEPU (28)], *Serratia marcescens* [SERMA (29)], *Salmonella typhimurium* [SALTY (29)], *Escherichia coli* [ECOLI (30)], *Proteus mirabilis* [PROMI (31)], *Mycoplasma capricolum* [MYCCA (32)], *Actinobacillus pleuropneumoniae* [ACTPAL (33)].

Hydrogen bonds. Hydrogen bonds between the two antiparallel β sheets may contribute to stabilization of the interface in both β and PCNA. The difference, however, is that within the PCNA interface there are eight potential hydrogen bonds between the sheets compared to only four in the β subunit. Perhaps the greater entropic force against stable association of a trimer versus a dimer requires more interactions among the interfaces of the protomers.

Hydrophobic forces. Hydrophobic interactions may also be important in the strength of the interface. The PCNA interface contains four pairs of hydrophobic amino acids which form a hydrophobic core (Fig. 4). In β, however, the interface hydrophobic core consists of only two pairs of hydrophobic residues (Fig. 4). This may suggest that hydrophobic interactions play an important role in the stabilization of a trimeric structure and again support the notion that more interactions are needed to stabilize a trimer relative to a dimer.

In β the two hydrophobic pairs are: Phe_{106} to Leu_{273} and Leu_{108} to Ile_{272} [(5); Fig. 4]. The position of the four hydrophobic residues of the interface have been highly conserved in all the bacterial species examined (Fig. 2). For example, Phe_{106} and Leu_{108} have been conserved in every species except *B.aphidicola* (Phe is changed to Tyr) and *M.capricolum* (Leu is changed to Ile). The Ile_{272} and Leu_{273} are conserved, though not identical and are replaced by either LeuVal or LeuLeu.

In PCNA the four hydrophobic pairs are: Ile_{78} to Leu_{154}, Ala_{112} to Val_{180}, Tyr_{114} to Ile_{181} and Leu_{116} to Ile_{182} (12). The eight hydrophobic residues in PCNA are not as conserved as in the prokaryotes. In the head interface the four hydrophobic amino acids in different species range from being identical to those in

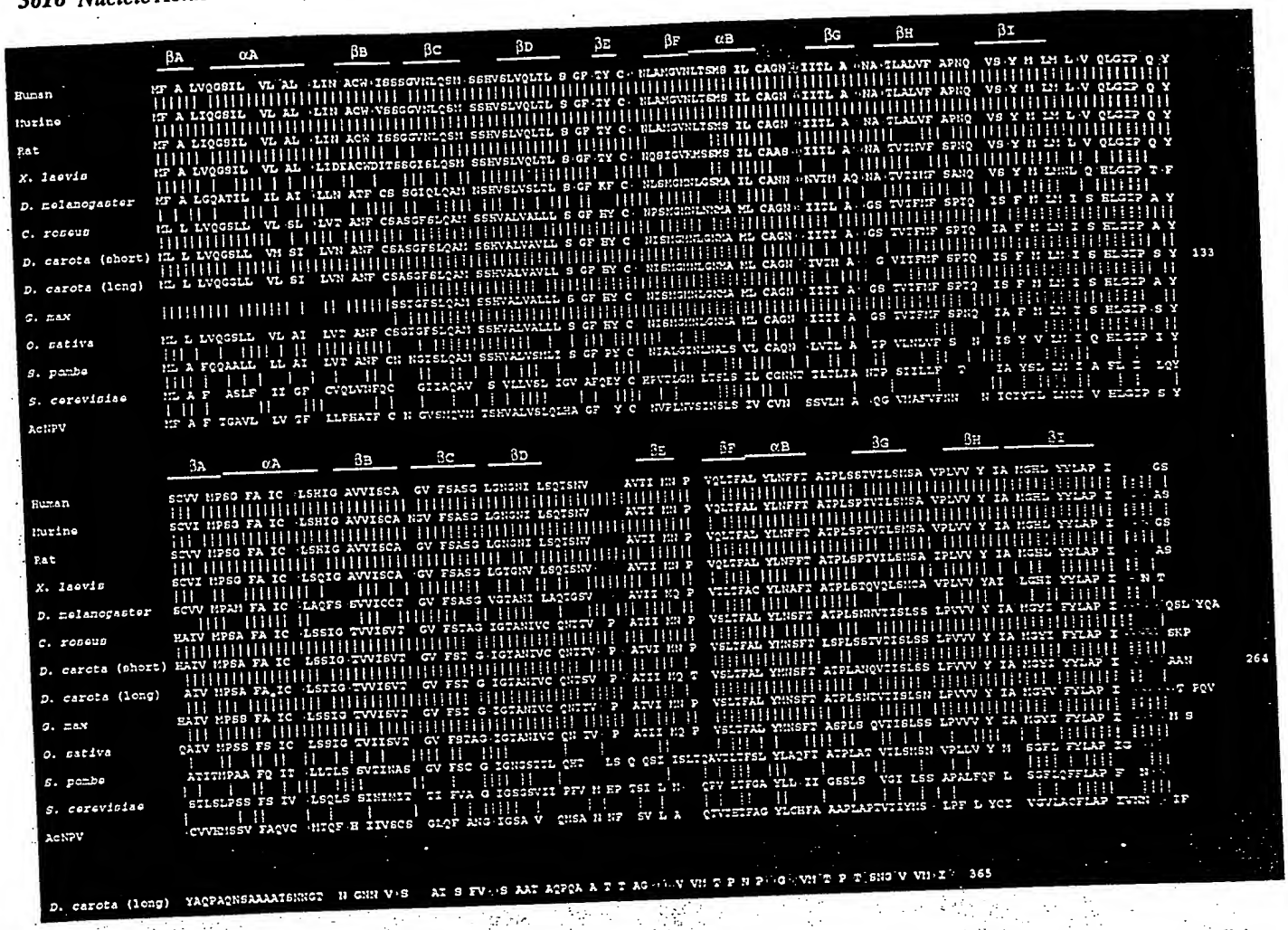


Figure 3. Sequence alignment among PCNA from different eukaryotes. The proteins are aligned as described in the legend to Figure 2. The sequences are split into two parts which correspond to the two globular domains of *S.cerevisiae* PCNA [except for the putative third domain for the amino acid sequence of *D.carrota* (long)]. Glu and Asp are red and Lys and Arg are blue. The secondary structures corresponding to the *S.cerevisiae* PCNA are indicated above each domain. To the right is the amino acid number at the end of each domain of the *D.carrota* (long) protein. The sequences are from the following references: human (34), murine (35), rat (36), *Xenopus laevis* [*X.laevis* (22)], *Drosophila melanogaster* [*D.melanogaster* (37)], *Catharanthus roseus* [*C.roseus* (38)], *Daucus carrota* [*D.carrota* (21)], *Glycine max* [*G.max* (39)], *Oriza sativa* [*O.sativa* (39)], *Saccharomyces pombe* [*S.pombe* (40)], *Saccharomyces cerevisiae* [*S.cerevisiae* (41)], *Autographa californica* nuclear polyhedrosis virus [AcNPV (14)].

S.cerevisiae or have been replaced by other hydrophobic amino acids. In the tail interface the picture is more complex. In all species the hydrophobic residue at position 154 is conserved (either Leu, Ile or Phe). However, in several species the amino acids in the triplet ValIleIle (residues 180–183) (Fig. 3) are replaced by either polar amino acids (Thr or Cys) or even a charged residue (Lys) in mammalian and *D.melanogaster* PCNAs (Fig 3).

Ion pairs. Ionic interactions between amino acids with opposing charges may also contribute to the stability of the interface. There are six putative ion pairs in the interface of *E.coli* β, but only one in yeast PCNA (Fig. 4). Interestingly, this is the opposite situation than that of the putative hydrophobic forces in the interface where there are four pairs of hydrophobic amino acids in PCNA and only two in β. This suggests that ion pairs may play a significant role in the strength of the interface in prokaryotic clamps.

In the primary sequence of the β subunit, the tail interface is the most tightly focused region of negative charge; five of seven residues are glutamates (residues 298–304; Figs 2 and 4). The putative tail interfaces of the other bacterial species also show this focused region of negative charge. Conversely, the head interface contains the positively charged partners, but these residues are scattered over a longer region.

The β sequences from other bacteria each have residues in positions that correspond to at least three of the ion pairs of *E.coli* β and they have additional charged residues in their interface region which could possibly form other ion pairs. The only putative ion pair conserved among all prokaryotes studied is the Lys_{74}–Glu_{298} pair. The Arg_{105}–Glu_{301} pair is reversed in *B.subtilis* (Glu_{105}–Lys_{301}). Residues corresponding to the buried Arg_{96}–Glu_{300} of *E.coli* β is present in all the species except for *B.subtilis* where both members of the pair are missing, suggesting intolerance at burying a lone member of this pair. In PCNA from all 13 species studied, there is an Asp at position 150 and either an Arg or Lys at position 110.

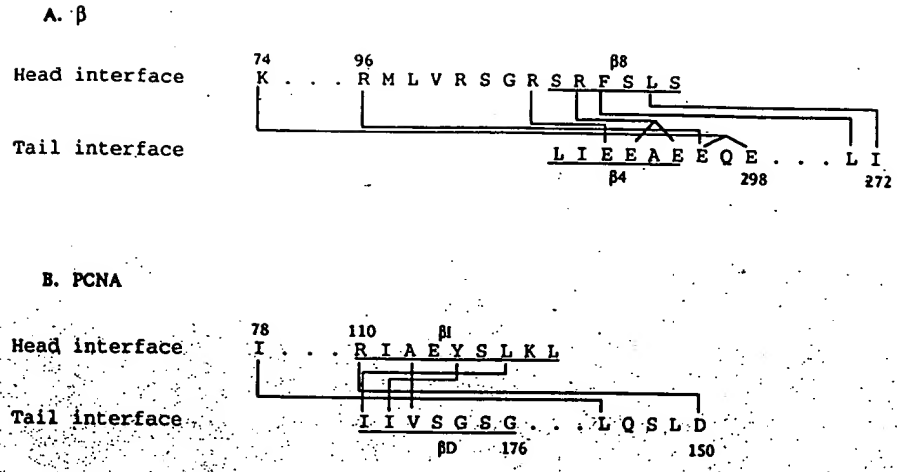


Figure 4. Residues that comprise the dimer interface of *E.coli* β and of *S.cerevisiae* PCNA. (A) The interface of the *E.coli* β and (B) the interface of the yeast PCNA. Amino acids that form the sheet structure at the dimer interface are underlined. Lines drawn between resides of the two interfaces indicate putative interactions which stabilize the structure. The N-terminal domain interface (head interface) is shown above the C-terminal domain interface (tail interface).

Table 1. Molecular mass, net charge and calculated pI of β subunits from different species (using the DNASTAR program)

Species	Amino acids	MW (Da)	Net charge (pH = 7)	pI
E.coli	366	40589.9	−9.982	5.218
B.subtilis	378	42106.6	−14.388	4.819
S.coelicoior	376	39959.4	−18.877	4.487
P.putida	366	40593.8	−10.279	5.095
P.mirabilis	367	40751.3	−9.286	5.137
B.aphidicola	365	42057.2	8.207	9.063

Only β subunits for which the entire sequence is known are listed.

Table 2. Molecular mass, net charge and calculated pI of PCNA from different species (using the DNASTAR program)

Species	Amino acids	MW (Da)	Net charge (pH = 7)	pI
Human	262	28899.5	−16.728	4.457
Murine	261	28787.5	−15.563	4.553
Rat	261	28751.4	−16.728	4.457
X.laevis	261	28899.3	−17.896	4.363
D.melanogaster	260	28832.5	−13.827	4.543
C.roseus	268	29767.6	−17.531	4.460
D.carrota (short)	264	29226.1	−15.532	4.556
D.carrota (long)	365	40105.1	−28.670	4.343
O.sativa	263	29274.2	−15.697	4.503
S.cerevisiae	258	28919.3	−19.837	4.261
S.pombe	260	28971.5	−17.671	4.367
AcNPV	256	28637.5	−9.072	5.219

Only PCNA for which the entire sequence is known are listed.

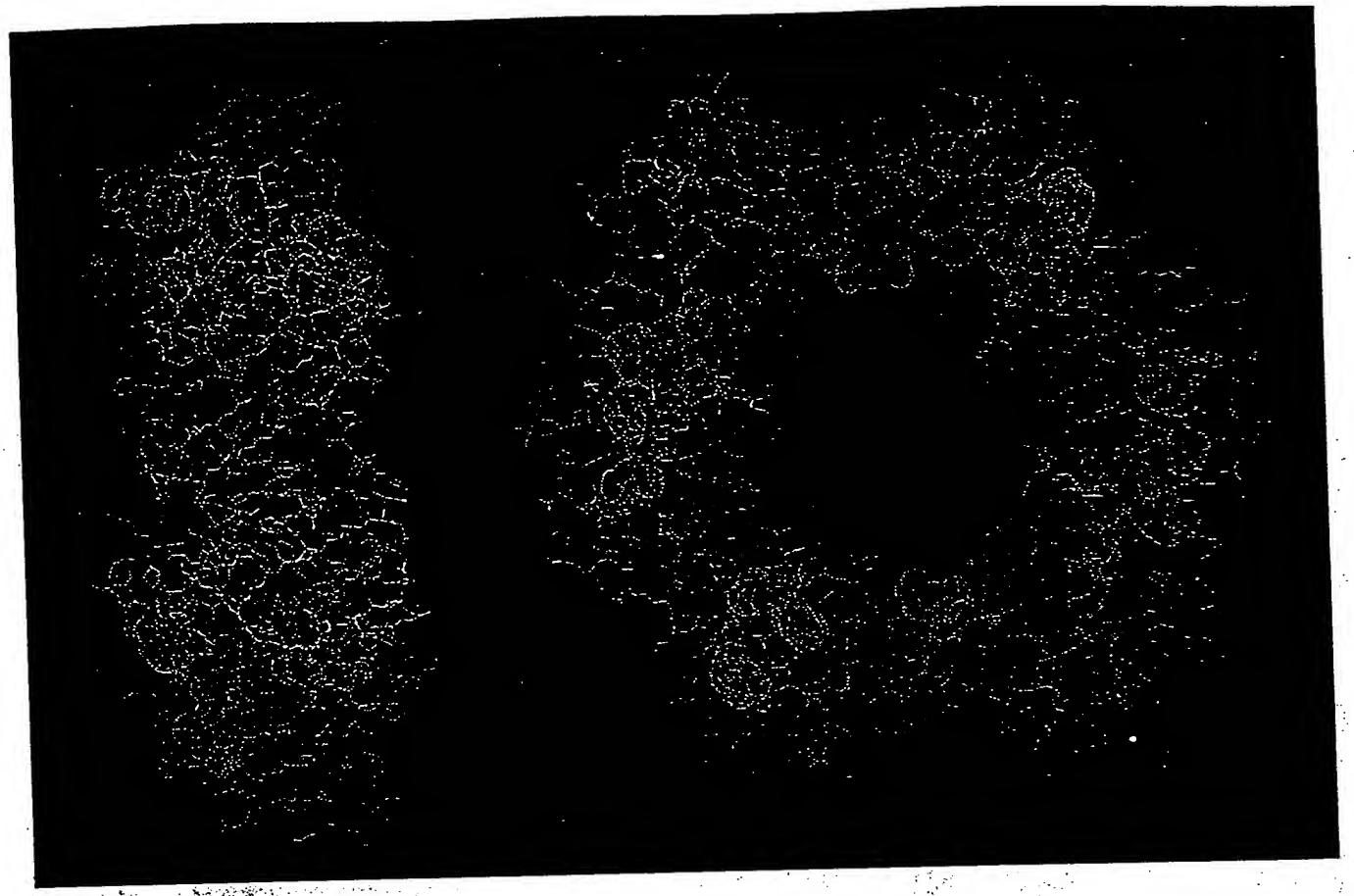


Figure 5. Highly conserved amino acids among β subunits of different species. The conserved amino acids (from Fig. 2) in loops are shown in yellow and those located in the α helices are shown in green.

The evolutionary conservation of charged pairs in these interfaces suggest they are important to the function of the clamps. The most obvious role is to stabilize the dimer interface. However, it is possible that the need to exclude water from the buried charges could require energy and thus destabilize the interface, adding to a balance between stability and instability in this region that may be needed to open and close the interface. Another possible role for charged residues at the interface derives from the high degree of symmetry in the ring. Computer modeling of a head to head β dimer results in a ring which looks like the head to tail ring, having tight interfaces with an antiparallel sheet as in the head to tail interface. Therefore, another possible function of the charged residues may be to provide specificity to form the head to tail dimer rather than the head to head dimer.

The interface may be a 'busy' region of protein interaction as this is the site at which work may be performed by the clamp loader to open and close the ring. Hence, another possible role of the charged residues at the interface may be to function in the recognition of other subunits of the polymerase. Also, if the polymerase (or another subunit) were to bind the clamp by spanning the interface it could act as a brace to further stabilize the interface thereby increasing the stability of the clamp on DNA and leading to greater processivity.

Sliding clamp interactions with other proteins

The sliding clamps must interact with at least two components of their corresponding DNA polymerase holoenzyme, the clamp loader and the DNA polymerase. However, these sliding clamps also interact with cellular components involved in other processes (Table 3). In the T4 system, gene 45 protein interacts with RNA polymerase to activate late gene transcription (13). Similarly, the PCNA homologue from the baculovirus AcNPV is important for late gene transcription (14). Beside its interactions with Pol III holoenzyme, the β subunit also interacts with DNA polymerase II an enzyme implicated in DNA repair. Human PCNA interacts with D-type cyclins (15), the cell cycle dependent kinase (CDK) inhibitor p21 (16,17) and the DNA damage induced gene, Gadd45 (18) (Table 3). *In vitro* replication studies have shown that the binding of p21 to PCNA inhibits replication (16,17). Gadd45, like p21, is induced by p53 upon DNA damage, presumably to block DNA replication. Consistent with this notion both p21 (16,17) and

Gadd45 (Jerard Hurwitz, pers. comm.) have been shown to inhibit DNA replication *in vitro*.

Table 3. Multiple enzymes interact with sliding clamps of prokaryotic and eukaryotic DNA polymerases

Clamp	Interacts with:
E.coli β	DNA polymerase II
	DNA polymerase III
T4 gene 45 protein	T4 DNA polymerase (gene 43 protein)
	E.coli RNA polymerase (modified by gene 33 protein and gene 55 protein)
Human PCNA	DNA polymerase δ
	DNA polymerase ε
	D-type cyclins
	p21 CDK inhibitor
	Gadd45 protein

The sequences of the different β subunits in Figure 2 are much more divergent than the PCNA genes in Figure 3, due to the greater evolutionary distance between the prokaryotic organisms relative to the eukaryotes. Hence, identical residues in the β subunits from all species may be expected to be functionally important. Among the β subunits analyzed only 15 residues are identical in all species (Fig. 2). These conserved amino acids are either in loops or in the α helices lining the central cavity (Fig. 5). Interestingly, the conserved amino acids on the loops are all on only one face of the ring. The conserved residues are good candidates for interactions with other proteins. Consistent with this, the C-termini of β which protrude from this face have been shown to be a touch point for interaction of β with both the clamp loader and the DNA polymerase, consistent with the recognition of β by both the polymerase and the clamp loader (V. Naktinis and M. O'Donnell, unpublished observation).

The C-terminal amino acids of the *E.coli* β subunit are important in the binding of both the γ complex and the polymerase (V. Naktinis and M. O'Donnell, unpublished observation). Examination of the C-terminus of PCNA shows a notable high negative charge. In all species there are 3–6 Glu or Asp residues adjacent to the C-terminus (Fig. 3). The three dimensional structure of the yeast PCNA shows that the three acidic amino acids form a 'hook' with the side chains pointing to the solvent (12). This highly acidic region is unique to PCNA and is not present in β or gene 45 protein. It is tempting to speculate that the acidic C-terminus of PCNA, lacking in β, may be involved in interaction with cellular regulators.

In the region preceding the acidic C-terminus of PCNA there are several sequence similarities to β. In all PCNAs, there is a conserved Lys before the Glu/Asp stretch and in all β subunits there is a positively charged amino acid in the C-terminus (Fig. 2). Additionally, both PCNA and β have a hydrophobic residue (Ile or Leu) following the positively charged amino acid and both have one Pro within the terminal four amino acids. The amino acid sequence similarities at the C-terminal region of both β and PCNA, together with the observation that the subunits of the *E.coli* γ complex share sequence similarities with the RF-C complex (19), suggest that this region in PCNA may be responsible for interaction with RF-C and/or with Pol δ.

Dimer and trimer sliding clamps are found in both prokaryotes and eukaryotes

All of the prokaryotic *dnaN* genes examined here encode proteins of approximately the same size (length of 365–378 residues; Table 1). Hence, they presumably form dimers like *E.coli* β. The 13 PCNA genes encode proteins that are ~2/3 the size of β (lengths of 257–269 residues; Table 2) and presumably form trimers like yeast PCNA.

Why did eukaryotes evolve trimers? The identical chain topology of the domains suggests that both PCNA and β evolved from a single domain that either formed a ring by hexamerising or that fused. The three domains of a β monomer required two fusions; the two domains of a PCNA monomer required only one fusion. Hence, it is possible that eukaryotes (and phage-T4) have yet to undergo the second fusion to form the three domain monomer like prokaryotes. Alternatively, the trimeric form may have evolved from the prokaryote three domain protein by deletion of one domain.

A trimeric ring composed of three protomers may not be as stable as a ring composed of two protomers, having a greater probability for protomer disassembly and a lower probability for three monomers to associate relative to only two. It is conceivable that a less stable ring is used to an advantage during the replication of the lagging strand. The lagging strand is synthesized as a series of fragments. Although the clamps exist in excess over the polymerase inside the cell, there are not enough clamps to provide one for each Okazaki fragment. Hence, the clamps must be recycled. The β subunit is very stable on DNA and a clamp unloading activity was identified to maintain a pool of available clamps within the cell (20). A less stable clamp on DNA may dissociate more easily from the DNA when the polymerase abandons it and thus a specific unloading mechanism may not be needed to maintain an intracellular pool of clamps.

It appears that both prokaryotes and eukaryotes make use of both type of rings, those composed of a trimer and those composed of a dimer. In the case of *E.coli*, it was recently demonstrated that a shorter version of β (called β*) is induced upon exposure of cells to UV light. β* is expressed from an internal promoter within the *dnaN* gene. This β* is comprised of the C-terminal 2/3 of normal β and hence each monomer contains two domains instead of three. Characterization of β* showed it behaves as a trimer on gel-filtration, presumably forming a 6-domain ring and it stimulates DNA synthesis by *Pol*III (Z. Livneh, pers. comm.).

A possible dimeric ring in eukaryotes was observed in an embryonic form of PCNA from *D.carrota* (21). Two distinct PCNA genes were isolated from *D.carrota*; one encodes a PCNA of typical length for a trimer (264 residues, 29.2 kDa) and the other encodes a larger protein of 365 amino acids (40.1 kDa). This is the size of a typical prokaryotic *dnaN* gene suggesting that the embryonic long form of PCNA may act as a dimer. Further, during oogenesis of the amphibian, *Xenopus laevis*, the presence of a 43 kDa protein that cross-reacts with anti-PCNA antibodies correlates with production of two differently sized transcripts (22).

Thus it appears that eukaryotes and prokaryotes alike may utilize rings made of dimers and trimers. Speculation as to why the chromosomal replicase of prokaryotes use the dimer form of

the clamp and eukaryotes use the trimer has been treated (23), but experimentally derived evidence pertaining to this question remains for further study.

ACKNOWLEDGEMENTS

We are grateful to Jerard Hurwitz and Zvi Livneh for sharing their data prior to publication. We also wish to thank Talluru Krishna and Jennifer Turner for their contributions to Figures 1 and 5. This work was supported by a grant from the National Institutes of Health (GM38839).

REFERENCES

1 Kuriyan, J. and O'Donnell, M. (1993) *J. Mol. Biol.*, **234**, 915–925.
2 Kelman, Z. and O'Donnell, M. (1995) *Annu. Rev. Biochem.*, **64**, 171–200.
3 Kelman, Z. and O'Donnell, M. (1994) *Curr. Op. Genet. Dev.*, **4**, 185–195.
4 Stukenberg, P.T., Studwell-Vaughan, P.S. and O'Donnell, M. (1991) *J. Biol. Chem.*, **266**, 11 328–11 334.
5 Kong, X.-P., Onrust, R., O'Donnell, M and Kuriyan, J. (1992) *Cell*, **69**, 425–437.
6 Gogol, E.P., Young, M.C., Kubasek, W.L., Jarvis, T.C. and von Hippel, P.H. (1992) *J. Mol. Biol.*, **224**, 395–412.
7 Reddy, M.K., Weitzel, S.E. and von Hippel, P.H. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 3211–3215.
8 Koboord, B.F. and Benkovic, S. (1995) *Curr. Biol.*, **5**, 149–157.
9 Sanders, G.M., Kassavetis, G.A. and Geiduschek, E.P. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 7703–7707.
10 Burgers, P.M. and Yoder E.L. (1993) *J. Biol. Chem.*, **268**, 19 923–19 926.
11 Tinker, R.L., Kassavetis, G.A. and Geiduschek, E.P. (1994) *EMBO J.*, **13**, 5330–5337.
12 Krishna, T.S.R., Kong, X-P., Gary, S., Burger, P.M. and Kuriyan, J. (1994) *Cell*, **79**, 1233–1243.
13 Herendeen, D.R., Kassavetis, G.A. and Geiduschek, E.P. (1992) *Science*, **256**, 1298–1303.
14 O'Reilly, D.R., Crawford. A.M. and Miller. L.K. (1989) *Nature*, **337**, 606.
15 Matsuka, S., Yamaguchi, M. and Matsukage, A. (1994) *J. Biol. Chem.*, **269**, 11 030–11 036.
16 Flores-Rozas, H., Kelman, Z., Dean, F., Pan, Z.-Q., Harper, J.W., Elledge, S.J., O'Donnell, M. and Hurwitz, J. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 8655–8659
17 Waga, S., Hannon, G.J., Beach, D. and Stillman, B. (1994) *Nature*, **369**, 574–578.
18 Smith, M.L., Chen, I.-T., Zhan, Q., Bae, I., Chen, C.-Y., Gilmer, T.M., Kastan, M.B., O'Conner, P.M. and Fornece, A.J. Jr (1994) *Science*, **266**, 1376–1380.
19 O'Donnell, M., Onrust, R., Dean, F.B., Chen, M. and Hurwitz, J. (1993) *Nucleic Acid Res.*, **21**, 1–3.
20 Stukenberg, P.T., Turner, J. and O'Donnell, M. (1994) *Cell*, **78**, 877–887.
21 Hata, S., Kouchi, H., Tanaka, Y., Minami, E., Matsumoto, T., Suzuka, I. and Hashimoto, J. (1992) *Eur. J. Biochem.*, **203**, 367–371.
22 Leibovici, M., Gusse, M., Bravo, R. and Méchali, M. (1990) *Dev. Biol.*, **141**, 183–192.
23 Kelman, Z. and O'Donnell, M. (1995) *Curr. Biol.*, **5**, 814.
24 Lai, C.-Y. and Baumann, P. (1992) *Gene*, **113**, 175–181.
25 Moriya, S., Ogasawara, N. and Yoshikawa, H. (1985) *Nucleic Acids Res.*, **7**, 2251–2265.
26 Fujita, M.Q., Yoshikawa, H. and Ogasawara, N. (1990) *Gene*, **93**, 73–78.
27 Calcutt, M.J. and Schmidt, F.J. (1992) *J. Bacteriol.*, **174**, 3220–3226.
28 Fujita, M.Q., Yoshikawa, H. and Ogasawara, N. (1989) *Mol. Gen. Genet.*, **215**, 381–387.
29 Skovgaard, O. and Hansen, F.G. (1987) *J. Bacteriol.*, **169**, 3976–3981.
30 Ohmori, H., Kimura, M., Nagata, T. and Sakakibara, Y. (1984) *Gene*, **28**, 159–170.
31 Skovgaard, O. (1990) *Gene*, **93**, 27–34.
32 Fujita, M.Q., Yoshikawa, H. and Ogasawara, N. (1992) *Gene*, **110**, 17–23.
33 Kroll, J.S. (unpublished) X63626.
34 Almendral, J. M., Huebsch, D., Blundell, P. A., Macdonald-Brown, H. and Bravo, R. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 1575–1579.
35 Yamaguchi, M., Hayashi, Y., Hirose, F., Matsuoka, S., Moriuchi, T., Shiroishi, T., Moriwaki, K. and Matsukage, A. (1991) *Nucleic Acids Res.*, **19**, 2403–2410.
36 Matsumoto, K., Moriuchi, T., Koji, T. and Nakane, P.K. (1987) *EMBO J.*, **6**, 637–642.
37 Yamaguchi, M., Nishida, Y., Noreuche, T., Hirose, F., Hui, F.F., Suzuki, Y. and Matsukaga, A. (1990) *Mol. Cell. Biol.*, **10**, 872–879.
38 Kodama, H., Ito, M., Ohnoshi, N., Suzuka, I. and Komamine, A. (1991) *Eur. J. Biochem.*, **197**, 495–503.
39 Suzuka, I., Hata, S., Matsuoka, M., Kosugi, S. and Hashimoto, J. (1991) *Eur. J. Biochem.*, **195**, 571–575.
40 Waseem, N.H., Labib, K., Nurse, P. and Lane, D.P. (1992) *EMBO J.*, **11**, 5111–5120.
41 Bauer, G.A. and Burgers, P.M. (1990) *Nucleic Acids Res.*, **18**, 261–265.

Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 1023–1027, February 1990
Biochemistry

Functions of replication factor C and proliferating-cell nuclear antigen: Functional similarity of DNA polymerase accessory proteins from human cells and bacteriophage T4

(simian virus 40 DNA replication/DNA binding/ATPase/evolution)

TOSHIKI TSURIMOTO AND BRUCE STILLMAN

Cold Spring Harbor Laboratory, P.O. Box 100, Cold Spring Harbor, NY 11724

Communicated by Bruce M. Alberts, November 20, 1989

ABSTRACT The proliferating-cell nuclear antigen (PCNA) and the replication factors A and C (RF-A and RF-C) are cellular proteins essential for complete elongation of DNA during synthesis from the simian virus 40 origin of DNA replication *in vitro*. All three cooperate to stimulate processive DNA synthesis by DNA polymerase δ on a primed single-stranded M13 template DNA and as such can be categorized as DNA polymerase accessory proteins. Biochemical analyses with highly purified RF-C and PCNA have demonstrated functions that are completely analogous to the functions of bacteriophage T4 DNA polymerase accessory proteins. A primer–template-specific DNA binding activity and a DNA-dependent ATPase activity copurified with the multisubunit protein RF-C and are similar to the functions of the phage T4 gene *44/62* protein complex. Furthermore, PCNA stimulated the RF-C ATPase activity and is, therefore, analogous to the phage T4 gene *45* protein, which stimulates the ATPase function of the gene *44/62* protein complex. Indeed, some primary sequence similarities between human PCNA and the phage T4 gene *45* protein could be detected. These results demonstrate a striking conservation of the DNA replication apparatus in human cells and bacteriophage T4.

At a DNA replication fork, appropriate assembly of accessory proteins and DNA polymerase (pol) is required to form an active holoenzyme (1, 2). Little was known about either the mechanism of eukaryotic DNA replication or the accessory proteins that interact with eukaryotic pols; however, recent studies on simian virus 40 (SV40) DNA replication *in vitro* have provided insights into DNA replication in mammalian cells (3). Although DNA polymerase α (pol α) was long thought to be a primary replicative pol in eukaryotes, the involvement of DNA polymerase δ (pol δ) in DNA replication was suggested only recently by studies with the SV40 system (4, 5) and has been demonstrated by genetic studies in yeast (6, 7). These studies suggest that a dimeric pol complex containing pol α, pol δ, and accessory proteins is responsible for coordinated replication of both the leading and lagging strands at the replication fork (3, 8). Pol δ contains a 3′–5′ exonuclease activity and can synthesize DNA processively on poly(dA)-oligo(dT) primer–template DNA in the presence of one accessory protein, proliferating-cell nuclear antigen (PCNA) (9, 10). Further studies demonstrated that pol δ can synthesize DNA processively and efficiently on a primed M13 single-stranded DNA (ssDNA) only in the presence of three replication factors, replication factor A (RF-A), PCNA, and replication factor C (RF-C) (8). RF-A is a multisubunit ssDNA binding protein that is required for both initiation and elongation of DNA replication *in vitro* and functions as a stimulatory factor for pol α and pol δ (8, 11–15). This latter function may be analogous to the activities of a number of viral, phage, and bacterial ssDNA binding proteins that stimulate their homologous pols (2, 3, 16). RF-C is required, along with PCNA, only for the elongation stage of SV40 DNA replication and appears to be required for the coordinated synthesis of leading and lagging strands (5, 17). Several properties of RF-C, for example, its function as a pol accessory protein and its moderate affinity for ssDNA bound to cellulose (17), suggested to us that it may be involved in template–primer recognition or as a molecular clamp holding the pol onto the template DNA. Analogous functions have been suggested for prokaryotic pol accessory proteins (1, 2, 16).

Studies on *Escherichia coli* and its phages have elucidated the general mechanisms for the initiation and elongation of DNA replication. Particularly relevant to this discussion are the studies on the bacteriophage T4 DNA replication proteins required at the replication fork [for review, see Cha and Alberts (18)]. The phage pol, encoded by gene *43*, synthesizes the leading and lagging strands at a replication fork, forming a dimeric pol complex. Pol accessory proteins, encoded by genes *44* and *62*, function as a DNA-dependent ATPase and primer-recognition protein complex. This DNA-dependent ATPase activity is stimulated by another protein, encoded by gene *45*, and together the gene *44/62* and gene *45* proteins complex and cooperate to stimulate the processivity of pol, forming a pol holoenzyme. In addition to these proteins, the helix-destabilizing protein (gene *32*, ssDNA binding protein) and the "primosome" proteins, encoded by genes *41* and *61* (DNA helicase and primase), also function at the replication fork to unwind the parental duplex DNA, form primers for Okazaki fragment synthesis on the lagging strand, and augment pol function. We have further characterized two human cell replication factors, RF-C and PCNA, and demonstrate that they are strikingly similar to proteins encoded by the phage T4 genes *44/62* and *45*, respectively.

MATERIALS AND METHODS

Replication Factors, RF-A, PCNA, and RF-C, and Pol δ. RF-A was purified from a human 293 cell cytoplasmic extract as described (15) and a ssDNA-cellulose fraction (650 μg/ml) was used. Two sources of PCNA were used in this study. PCNA (200 μg/ml) purified from human 293 cells by a published procedure (4) was used in Figs. 2 and 4. PCNA that had the same specific activity as PCNA from human cells was produced in *E. coli* harboring a plasmid carrying the human PCNA cDNA sequence (19) under the control of bacteriophage T7 promoter. The *E. coli*-produced PCNA was purified

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: pol, DNA polymerase; pol α, DNA polymerase α; pol δ, DNA polymerase δ; pol III, DNA polymerase III; RF-A, replication factor A; RF-C, replication factor C; PCNA, proliferating-cell nuclear antigen; SV40, simian virus 40; ATP[S], adenosine 5′-[γ-thio]triphosphate; ssDNA, single-stranded DNA.

to homogeneity by four steps (K. Fien and B.S., unpublished results) and used in Fig. 3. RF-C was purified from human 293 cell nuclear extracts through four steps (ssDNA-cellulose fraction; 60 or 40 μg of RF-C per ml) or five steps (glycerol gradient fraction; 8 μg of RF-C per ml) as published (17). Pol δ was purified from calf thymus (90 g) by five steps as described (8, 20) and the ssDNA-cellulose fraction with a specific activity of 6.8×10^3 units/mg was used. One unit of pol activity was defined as the incorporation of 1 nmol of dTMP at 37°C in 1 hr under conditions described in ref. 8.

DNA Synthesis Reaction on a Primed M13 ssDNA with Pol δ. The reaction mixture (25 μl) containing 30 mM Hepes (pH 7.8), 30 mM NaCl, 7 mM $MgCl_2$, 0.5 mM dithiothreitol, bovine serum albumin (0.1 mg/ml), all four dNTPs (each at 0.05 mM) with [α-^{32}P]dATP (2000 cpm/pmol), 100 ng of M13 ssDNA (M13mp18) primed with a 3-fold molar excess of a unique 17-base sequencing primer (primer 1211 from New England Biolabs), 200 ng of PCNA, 1 μg of RF-A and 0.27–0.54 unit of pol δ was incubated at 37°C for 30 min, and acid-insoluble radioactivity was determined.

ATPase Assay. The assay for RF-C ATPase was essentially the same as described (21). The reaction mixture (25 μl) containing 50 mM of Tris·HCl (pH 7.9), 0.1 M NaCl, 1 mM dithiothreitol, 2 mM $MgCl_2$, bovine serum albumin (100 μg/ml), 0.1 mM [γ-^{32}P]ATP, and the indicated DNAs was incubated at 37°C for 30 min.

RESULTS

RF-C Binds Specifically to a Primer–Template DNA. RF-C and PCNA cooperate with RF-A to stimulate the putative leading-strand pol, pol δ (8). As noted above, this suggested that they may be similar to prokaryotic pol accessory proteins. Since RF-C bound to DNA, the DNA binding activity of RF-C was studied by nitrocellulose filter binding assays with ^{32}P-labeled poly(dA) [a (template) ssDNA] or poly(dA)·oligo(dT) (a primer–template DNA). As shown in Fig. 1*A*, RF-C bound to poly(dA)·oligo(dT) but not to poly(dA). The binding specificity of RF-C was demonstrated by testing the ability of various unlabeled DNAs to act as competitors for RF-C binding to ^{32}P-labeled poly(dA)·oligo(dT) (Fig. 1*B*). Under these conditions, RF-C bound to ssDNA carrying primers but not to ssDNAs or double-stranded DNAs or to an RNA·DNA template–primer [poly(A)·oligo(dT)]. This primer–template-specific DNA binding activity cosedimented in a glycerol gradient with two other activities of RF-C; the ability to stimulate SV40 DNA replication *in vitro* and the stimulation of pol δ on a primed M13 ssDNA in the presence of RF-A and PCNA (Fig. 2). As reported (17), the latter two activities cosedimented with a multisubunit protein containing polypeptides with apparent molecular masses of 140, 41, and 37 kDa. This analysis suggests that the primer–template binding activity is due to RF-C and is not a contaminant in the purified preparation.

RF-C Is an ATPase Stimulated by a Primer–Template DNA and PCNA. Several functions for prokaryotic pol accessory proteins have been reported (2, 23–28) and, among these, the well-characterized bacteriophage T4 gene *44/62* protein complex has a specific primer–template DNA binding activity (23, 25). Moreover, this protein complex also contains a DNA-dependent ATPase activity that is stimulated by the presence of 3′ ends on the DNA (27, 28). Thus, we tested the possibility that RF-C might have a similar ATPase activity. Fractions from the glycerol gradient that exhibited the specific primer–template binding activity also contained an ATPase activity when assayed in the presence of poly(dA)·oligo(dT) (Fig. 2*B*), indicating that RF-C also functions as an ATPase. As shown in Fig. 3*A*, highly purified RF-C has a low level of DNA-independent ATPase activity, but this activity is stimulated severalfold by either ssDNAs or double-stranded DNAs

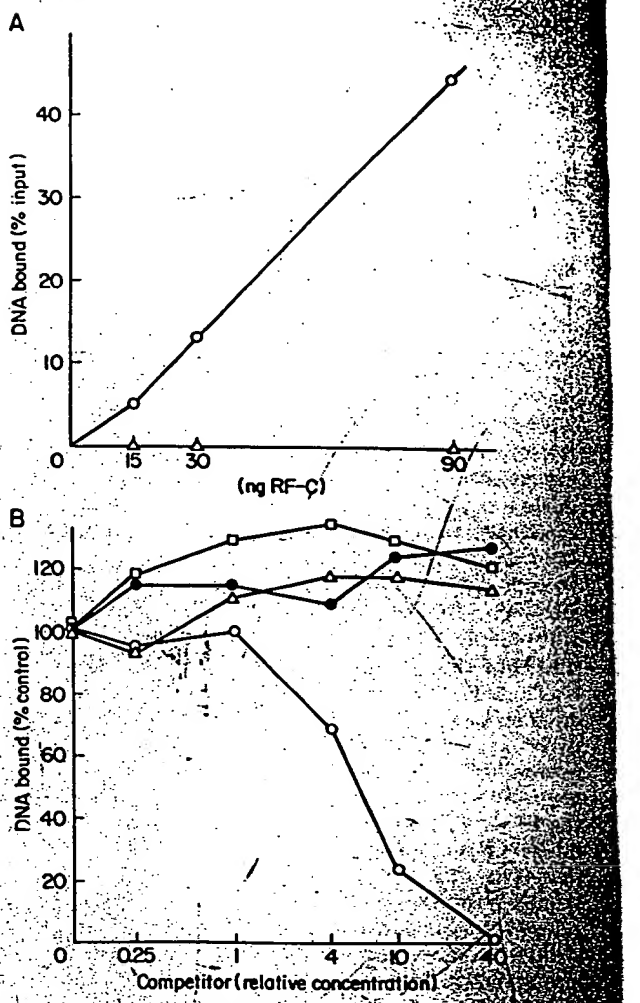


FIG. 1. DNA binding properties of RF-C. (*A*) Either ^{32}P-labeled poly(dA) (5 ng per reaction mixture, Δ) or ^{32}P-labeled poly(dA)·oligo(dT)$_{12}$ (1:4 molar ratio, 5 ng per reaction mixture, ○) were incubated with various amounts of RF-C (ssDNA-cellulose fraction, 60 μg/ml) at 0°C in a 25-μl reaction mixture containing the same buffer as used in the pol δ assay, but with 40 mM NaCl for 30 min. The mixture was then filtered through alkali-washed nitrocellulose (22), the protein-bound DNA was trapped on the filter, and the radioactivity was measured. The 100% value of input cpm was 3000 cpm. (*B*) Competition of the binding by various DNAs or RNA was tested. All of the reaction mixtures contained ^{32}P-labeled poly(dA)·oligo(dT) and 60 ng of RF-C as shown in *A*. The mixtures were incubated with the indicated molar amounts (based on nucleotide) of competitors at 0°C for 30 min and then filtered on nitrocellulose. The 100% value of the control with no competitor corresponds to 20% binding of the input label. The competitors were poly(dA) (Δ), poly(dA)·oligo(dT) (○), poly(A)·oligo(dT) (●, 4:1 molar ratio), or double-strand adenovirus type 2 DNA digested with *Hin*dIII (□).

[poly(dA) or adenovirus DNA digested with *Hin*dIII, respectively] and to a greater extent by the presence of primers on the ssDNA [poly(dA)·oligo(dT)]. Therefore, RF-C is a DNA-dependent ATPase but is not absolutely dependent upon primers bound to the template DNA. A strikingly similar dependence on DNA has been observed for the ATPase activity of the phage T4 genes *44/62* protein complex (27).

In addition to the stimulatory effect of DNAs, the phage T4 gene *44/62* protein complex ATPase activity is also stimulated by another accessory protein, the gene *45*-encoded protein (26–28). Since the phage T4 gene *45* protein and PCNA do not bind directly to DNA but can form a complex with their respective pols (2, 9), we predicted that PCNA

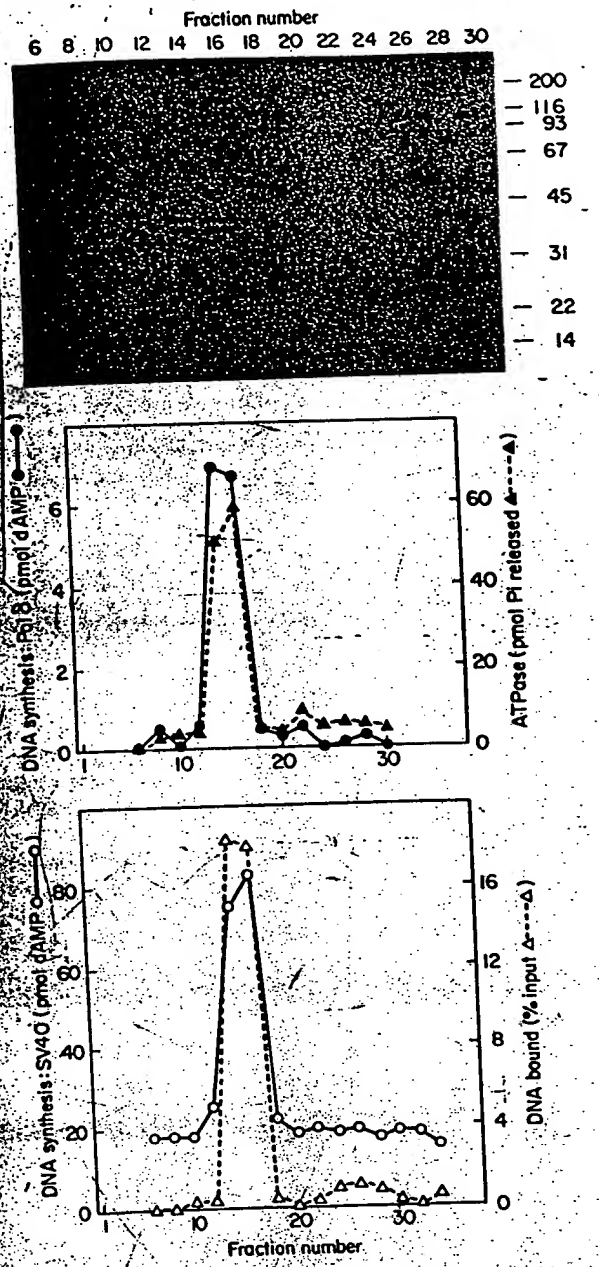


FIG. 2. Cosedimentation of primer–template DNA binding and DNA-dependent ATPase activities with RF-C in a glycerol gradient. A ssDNA-cellulose fraction of RF-C (100 μl; 40 μg/ml of protein) was loaded onto a 5-ml 15–35% (vol/vol) glycerol gradient in buffer A containing 0.1 M NaCl as described (17). After centrifugation in a SW 50.1 rotor at 49,000 rpm at 4°C for 24 hr, 38 fractions were collected from the bottom. The positions of marker proteins, which were run in a parallel gradient were catalase (11 S) at fraction 9, alcohol dehydrogenase (7.4 S) at fraction 16, bovine serum albumin (4.3 S) at fraction 24, ovalbumin (3.7 S) at fraction 26, and cytochrome C (1.9 S) at fraction 32. (*A*) A 10-μl aliquot of every second fraction was separated in an $NaDodSO_4$/polyacrylamide gel (12.5%) and proteins were stained with silver. Molecular mass markers (kDa) are shown on the right. (*B*) Stimulation of pol δ on a primed M13 ssDNA in the presence of RF-A and PCNA, and DNA-dependent ATPase activity measured with 26 μM poly(dA)-oligo(dT) (1:4 molar ratio), were determined using 4 μl of each fraction. DNA-independent ATPase activity was also measured in fractions 14 and 16 and was ≈10% of the level obtained with DNA present (data not shown). (*C*) Complementation assay for RF-C in SV40 DNA replication *in vitro* and DNA binding to a primer–template. The RF-C complementation was done in a 50-μl reaction mixture containing 350 μg of fraction I*, 1.7 μg of SV40 tumor antigen, 200 ng of topoisomerase I, 90 ng of topoisomerase II, 300 ng of pSVO10, and 8 μl of each fraction as described (15). After the incubation at 37°C for 1 hr, acid-insoluble cpm were measured. The primer–template DNA binding activity was measured as described in the legend to Fig. 1 with 3 μl of each fraction, but in the presence of 50 mM NaCl. DNA binding activity using poly(dA) was also tested in parallel with the same fractions; however, no DNA binding was detected (data not shown).

would be analogous to the phage T4 gene *45* protein and would stimulate the RF-C DNA-dependent ATPase. This was tested by addition of various amounts of purified PCNA into the RF-C ATPase reaction in the presence of primer–template DNA. The DNA-dependent ATPase activity of RF-C was stimulated up to 4-fold in the presence of saturating amounts of PCNA (Fig. 3*B*). In the absence of RF-C, highly purified PCNA revealed a background level of ATPase, so that it is possible that this minor ATPase observed in the PCNA preparation is greatly stimulated by interaction with RF-C. But the PCNA-stimulated ATPase activity displayed the same specificity for the various DNAs as the RF-C ATPase activity shown in Fig. 3*A* (data not shown), suggesting that most of the ATPase activity observed is intrinsic to the RF-C polypeptides. Therefore, this result demonstrates an interaction between RF-C and PCNA similar to that observed between the phage T4 gene *44/62* complex and the gene *45* protein.

Hydrolysis of ATP Is Required for Processive DNA Synthesis by Pol δ. In the phage T4 DNA replication system, hydrolysis of ATP or dATP by the accessory proteins is required for the formation of a stable initiation complex containing accessory proteins and pol bound to the 3′ end of the DNA primer (2, 23, 29). If RF-C and PCNA are functionally analogous to the phage T4 gene *44/62* protein complex and gene *45* protein, respectively, stimulation of pol δ activity on primed M13 ssDNA by RF-C and PCNA should require hydrolysis of ATP. As reported (8), DNA synthesis by pol δ on primed M13 ssDNA was stimulated about 2-fold by RF-C and PCNA compared to incorporation by pol δ alone. If 1 mM ATP was added to this reaction, the DNA synthesis was further stimulated about 3-fold (Fig. 4). The ATP-independent processive DNA synthesis by pol δ in the presence of RF-C and PCNA may have been due to the utilization of dATP instead of ATP, as has been reported in phage T4 system (2). When all three replication proteins, RF-A, RF-C, and PCNA, were added to reactions containing pol δ and primed M13 ssDNA, addition of ATP had no effect (Fig. 4). This is probably due to maximal stimulation of pol δ by the replication accessory proteins and dATP. We have demonstrated that under these conditions, DNA synthesis by pol δ is very processive (8). However, ATP (or dATP) hydrolysis is required for this stimulation of pol δ because the addition of 1 mM adenosine 5′-[γ-thio]triphosphate (ATP[S]), a non-hydrolyzable analogue of ATP, to the reaction completely abolished the stimulation of processive DNA synthesis by the replication accessory proteins (Fig. 4). ATP[S] did not inhibit the small amount of synthesis obtained with pol δ plus PCNA, suggesting that the effect of ATP[S] was RF-C-dependent and not the result of competitive inhibition of dAMP incorporation. Therefore, these results demonstrated that processive DNA synthesis by pol δ on primed M13 ssDNA requires three accessory proteins and hydrolysis of ATP (or dATP), as has been reported in phage T4 system.

DISCUSSION

These biochemical analyses clearly indicated that RF-C and PCNA are functionally analogous to the phage T4 gene *44/62* protein complex and the gene *45* protein, respectively. It was, therefore, of interest to determine whether there were any

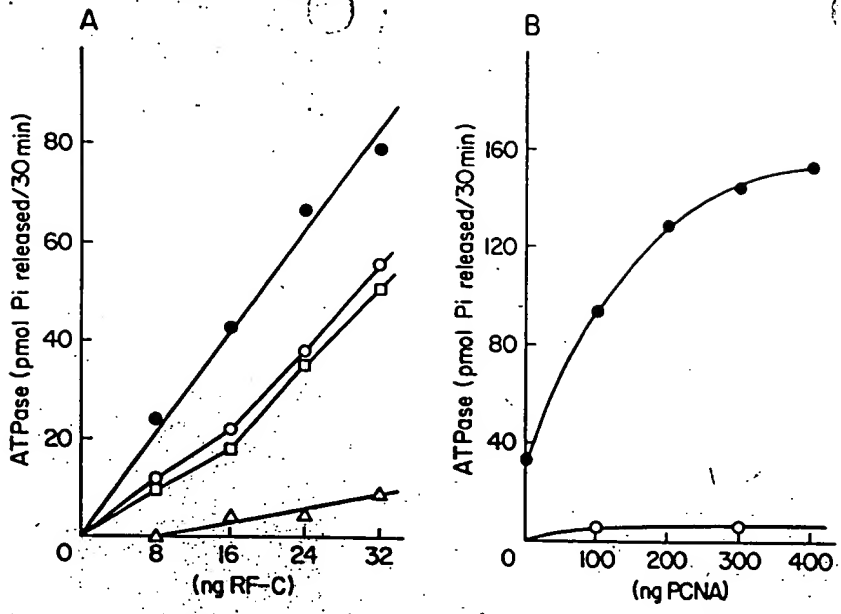


FIG. 3. ATPase activity of RF-C. (*A*) Titration of RF-C without DNA (△) or with 26 μM poly(dA) (□), 26 μM adenovirus DNA digested with *Hin*cII (○), or 26 μM poly(dA)-oligo(dT) (1:4 molar ratio) (●). (Note that all the DNA concentrations are expressed as μM of nucleotide.) A 25-μl reaction mixture containing the indicated amounts of RF-C was incubated with DNA and [γ-^{32}P]ATP and the released P_i from ATP was determined. (*B*) Addition of PCNA to RF-C ATPase assay. Increasing amounts of PCNA were incubated in the reaction mixture containing 26 μM poly(dA)-oligo(dT) with (●) and without (○) 24 ng of RF-C, and the amount of released P_i was measured.

structural similarities between them; however, only a comparison between human PCNA (19) and gene *45* protein (30) is possible at present. A computer search of their primary amino acid sequence revealed limited regional similarities (31–50% similarity) that could be aligned in a linear arrangement, albeit with several gaps in each amino acid sequence (Fig. 5). The fact that these identities span the length of the amino acid sequences and that these proteins are functionally equivalent suggests that the proteins are evolutionarily related and that the identical residues are important for function in both proteins.

The major replicative pol in *E. coli* (pol III) also associates with multiple subunits and requires hydrolysis of ATP for processive DNA synthesis (1, 24, 32). In this case, the $\gamma\delta\delta'$ complex and the β subunit of pol III holoenzyme seem to function similarly to RF-C and PCNA, respectively. The β subunit was proposed (9) to have a function similar to PCNA but the amino acid sequence deduced from its coding sequence, the *dnaN* gene, revealed little significant similarity with human PCNA (data not shown). Thus, there appears to be a particular functional and structural relationship between the phage T4 and mammalian cell replication components. Indeed, the T4 gene *45*-encoded protein stimulated the RF-C ATPase activity (unpublished results).

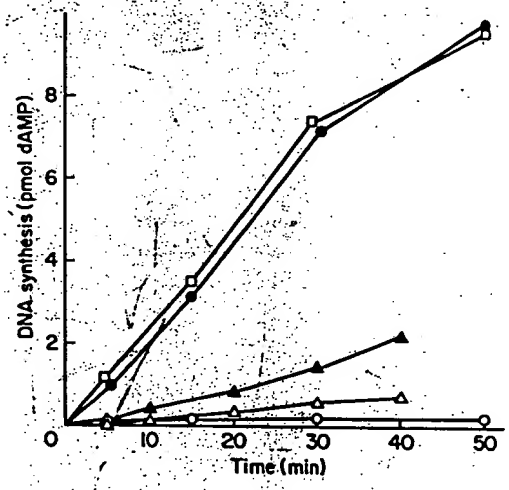


FIG. 4. Time-course of DNA synthesis with pol δ. The reaction mixture (25 μl) contained 0.54 unit of pol δ (8), 200 ng of PCNA, 40 ng of RF-C, and, where indicated, 650 ng of RF-A, 1 mM ATP or 1 mM ATP[S]. Samples (3 μl) were withdrawn at the indicated times and acid-insoluble radioactivity was measured. Values are shown as the incorporation of dAMP per 25 μl of the reaction mixture. Components used were PCNA and RF-C (△); PCNA, RF-C, and ATP (▲); PCNA, RF-C, and RF-A (□); PCNA, RF-C, RF-A, and ATP (●); PCNA, RF-C, RF-A, and ATP[S] (○). Experiments with RF-C and PCNA (with and without ATP) were done separately from other experiments. Experiments with RF-C, PCNA, and ATP[S] or pol δ alone were also done in parallel and yielded almost the same results shown by the open circles (data not shown).

As mentioned above, two pols, α and δ, are involved in eukaryotic DNA replication, and in the yeast *Saccharomyces cerevisiae* pols I and III correspond to pols α and δ, respectively (6, 7, 33). Genes coding for human pol α and yeast pols I and III have been isolated. Comparison of their amino acid sequences demonstrated the presence of six conserved regions that may correspond to several functional domains (7, 34). It is striking that bacteriophage T4 pol also belongs to this class of pols by sequence similarity (35). In contrast, the *E. coli* pol III has almost no similarity, suggesting that *E. coli* may have evolved further from the ancestral form. Thus it is likely that during development of eukaryotes an ancestral pol gene of the phage T4 type duplicated and diverged to yield the pol α and pol δ types. This study further suggests that the pol accessory proteins were also conserved during eukaryotic evolution; however, in eukaryotes, the accessory protein complex functions with pol δ, probably because it appears to be the processive leading-strand pol. On the other hand, pol α, which does not appear to interact with PCNA but does interact with RF-C and RF-A, has become partially independent of the accessory proteins (8). Interestingly, pol α has kept the interaction with a DNA primase, presumably because it functions as the lagging-strand pol. Chia and Alberts (36) have demonstrated that the phage T4 primase, encoded by gene *61*, is only required for lagging-strand DNA replication and not for leading-strand synthesis.

Fractionation of factors required for SV40 DNA replication *in vitro* has been a powerful method for identifying replication components from eukaryotic cells. Two of them were identified as accessory proteins for pol δ and are undoubtedly involved directly in cellular DNA replication; however, other replication components remain to be purified and characterized.

A remarkable similarity between bacteriophage T4 and eukaryotes has been noted. The group I introns present in some phage T4 mRNAs are related to the group I introns present in some eukaryote RNAs (37). Shub *et al.* (37)

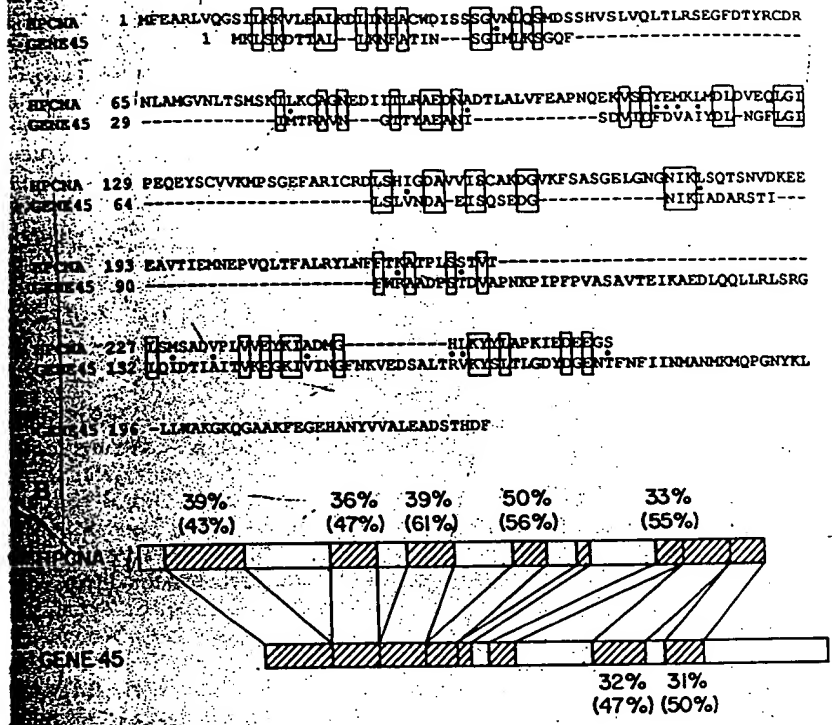


FIG. 5. Comparison of human PCNA (HPCNA) and phage T4 gene 45 protein (GENE45). (*A*) Amino acid sequences were obtained from refs. 19 and 30. Homologies between human PCNA and gene 45 protein were detected using the ALIGN program (31) and manual inspection. Identical amino acids are boxed and conserved amino acids are indicated with a dot. (*B*) Regions of homology between the two proteins are indicated and the numbers indicate percent identity and percent conservation (in parentheses) between them with respect to the human PCNA sequence. The single-letter amino acid code is used.

...(el that these group I intron similarities could be due ...rtical acquisition of an ancestral form present before the ...ergence of prokaryotes and eukaryotes or horizontal ...mission of introns after the divergence of prokaryotes ...eukaryotes. The observation that the DNA replication ...chinery in phage T4 and human cells are also similar favors ...rtical acquisition as the mechanism for the evolutionary ...atedness. Of course it is possible that phage T4 infected ...h prokaryotic and eukaryotic cells after their divergence. ...Although it is interesting that the replication apparatus is ...nly conserved between human cells and a bacteriophage ...t infects *E. coli*, ultimately we must also understand how ...eukaryotic replication apparatus replicates chromatin, ...just DNA, and it is this aspect of DNA replication that ...y make the eukaryotic replication apparatus unique.

...e thank John Diffley for advice, Richard Roberts and Winship ...r for valuable comments on the manuscript, Jack Barry and ...ce Alberts for phage T4 proteins, J. Duffy for the artwork, and ...Weinkauff for typing the manuscript. This work was supported by ...nt CA13106 from the National Institutes of Health.

...McHenry, C. S. (1988) *Annu. Rev. Biochem.* **57**, 519–550.
...Alberts, B. M., Barry, J., Bediger, P., Burke, R. L., Hibner, U., Liu, ...C.-C. & Sheridan, R. (1980) in *Mechanistic Studies of DNA Replication ...and Genetic Recombination*, ICN–UCLA Symposia on Molecular & ...Cellular Biology, ed. Alberts, B. M. (Academic, New York) Vol. 19, pp. ...449–471.
...Stillman, B. (1989) *Annu. Rev. Cell. Biol.* **5**, 197.
...Prelich, G., Kostura, M., Marshak, D. R., Mathews, M. B. & Stillman, ...B. (1987) *Nature (London)* **326**, 471–475.
...Prelich, G. & Stillman, B. (1988) *Cell* **53**, 117–126.
...Sidey, K. C., Budd, M. E. & Campbell, J. L. (1989) *Cell* **58**, 599–605.
...Boulet, A., Simon, M., Faye, G., Bauer, G. A. & Burgers, P. M. J. ...(1989) *EMBO J.* **8**, 1849–1854.
...Tsurimoto, T. & Stillman, B. (1989) *EMBO J.* **8**, 3883–3889.
...Tan, C. K., Castillo, C., So, A. G. & Downey, K. M. (1986) *J. Biol. ...Chem.* **261**, 12310–12316.
...Prelich, G., Tan, C. K., Kostura, M., Mathews, M. B., So, A. G., ...Downey, K. & Stillman, B. (1987) *Nature (London)* **326**, 517–520.
...Fairman, M. P. & Stillman, B. (1988) *EMBO J.* **7**, 1211–1218.
...Wobbe, C. R., Weissbach, L., Borowiec, J. A., Dean, F. B., Murakami, ...Y., Bullock, P. & Hurwitz, J. (1987) *Proc. Natl. Acad. Sci. USA* **84**, ...1834–1838.
13. Wold, M. S. & Kelly, T. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2523–2527.
14. Wold, M. S., Weinberg, D. H., Virsup, D. M., Li, J. J. & Kelly, T. J. (1989) *J. Biol. Chem.* **264**, 2801–2809.
15. Tsurimoto, T., Fairman, M. P. & Stillman, B. (1989) *Mol. Cell. Biol.* **9**, 3839–3849.
16. Kornberg, A. (1980) *DNA Replication* (Freeman, San Francisco).
17. Tsurimoto, T. & Stillman, B. (1989) *Mol. Cell. Biol.* **9**, 609–619.
18. Cha, T.-A. & Alberts, B. M. (1988) in *Cancer Cells 6, Eukaryotic DNA Replication*, eds. Kelly, T. & Stillman, B. (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 1–10.
19. Almendral, I. M., Huebsch, D., Blundell, P. A., MacDonald-Bravo, H. & Bravo, R. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 1575–1579.
20. Lee, M. Y. T., Tan, C.-K., Downey, K. M. & So, A. G. (1984) *Biochemistry* **23**, 1906–1913.
21. Wickner, S. & Hurwitz, J. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 921–925.
22. McEntee, K., Weinstock, G. M. & Lehman, I. R. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 857–861.
23. Huang, C.-C., Hearst, J. E. & Alberts, B. M. (1981) *J. Biol. Chem.* **256**, 4087–4094.
24. Maki, S. & Kornberg, A. (1988) *J. Biol. Chem.* **263**, 6561–6569.
25. Selick, H. E., Barry, J., Cha, T.-A., Munn, M., Nakanishi, M., Wong, M. L. & Alberts, B. M. (1987) in *DNA Replication and Recombination*, UCLA Symposia on Molecular & Cellular Biology, New Series, eds. McMacken, R. & Kelly, T. (Liss, New York) Vol. 47, pp. 183–214.
26. Rush, J., Lin, T.-C., Quinones, M., Spicer, E. K., Douglas, I., Williams, K. R. & Konigsberg, W. H. (1989) *J. Biol. Chem.* **264**, 10943–10953.
27. Mace, D. C. & Alberts, B. M. (1984) *J. Mol. Biol.* **177**, 279–293.
28. Jarvis, T. C., Paul, L. S., Hockensmith, J. W. & von Hippel, P. H. (1989) *J. Biol.Chem.* **264**, 12717–12729.
29. Mace, D. C. & Alberts, B. M. (1984) *J. Mol. Biol.* **177**, 313–327.
30. Spicer, E. K., Noble, J. A., Nossal, N. G., Konigsberg, W. H. & Williams, K. R. (1982) *J. Biol. Chem.* **257**, 8972–8979.
31. Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983) *Methods Enzymol.* **91**, 524–545.
32. O'Donnell, M. E. (1987) *J. Biol. Chem.* **262**, 16558–16565.
33. Bauer, G. A., Heller, H. M. & Burgers, P. M. J. (1988) *J. Biol. Chem.* **263**, 917–924.
34. Wong, S. W., Wahl, A. F., Yuanm, P.-M., Arai, N., Pearson, B. E., Arai, K.-I., Korn, D., Hunkapiller, M. W. & Wang, T. S. (1988) *EMBO J.* **7**, 37–47.
35. Spicer, E. K., Rush, J., Fung, C., Reha-Krantz, L. J., Karam, J. D. & Konigsberg, W. H. (1988) *J. Biol. Chem.* **263**, 7478–7486.
36. Cha, T.-A. & Alberts, B. M. (1989) *J. Biol. Chem.* **264**, 12220–12225.
37. Shub, D. A., Gott, J. M., Xu, M.-Q., Lang, B. F., Michel, F., Tomaschewski, J., Pedersen-Lane, J. & Belfort, M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 1151–1155.

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 268, No. 16, Issue of June 5, pp. 11758-11765, 1993
Printed in U.S.A.
© 1993 by The American Society for Biochemistry and Molecular Biology, Inc.

DNA Polymerase III Accessory Proteins

I. *holA* and *holB* ENCODING δ AND δ'*

(Received for publication, September 16, 1992, and in revised form, January 20, 1993)

Ziming Dong‡§, Rene Onrust§¶, Maijā Skangalis‡, and Mike O'Donnell‡¶‖

From the ‡Howard Hughes Medical Institute and the ¶Microbiology Department, Hearst Research Foundation, Cornell University Medical College, New York, New York 10021

The genes encoding the δ and δ' subunits of the 10-subunit *Escherichia coli* replicase, DNA polymerase III holoenzyme, have been identified and sequenced. The *holA* gene encoding δ is located downstream of *rlpB* at 15.2 min and predicts a 38.7 kda protein. The *holB* gene encoding δ' is located at 24.3 min and predicts a 36.9-kDa protein. Hence the δ and δ' subunits are unrelated proteins encoded by separate genes. The genes have been used to express and purify δ and δ' in quantity. The predicted amino acid sequence of δ' is homologous to the sequences of the τ and γ subunits revealing a large amount of structural redundancy within the holoenzyme.

DNA polymerase III holoenzyme, the replicase of *Escherichia coli*, is composed of 10 subunits (α, ε, θ, τ, γ, δ, δ', χ, ψ, β) (1). The holoenzyme[1] is fast (750 nucleotides(s)) and highly processive (>100 kb) in synthesis (1–4). This speed is approximately the observed rate of replication fork movement in *E. coli* (5), and the high processivity is consistent with the single origin from which two opposed replication forks travel around the 4-Mb circular chromosome. The three-subunit polIII core subassembly (α, DNA polymerase; ε, proofreading 3'–5' exonuclease; and θ) is slow (20 nucleotides(s)) and only processive for approximately 11 nucleotides (2). Speed and processivity are conferred onto the core polymerase by the γ complex (γ, δ, δ', χ, and ψ) and β accessory proteins (6–8). Biochemical and x-ray studies show that β is shaped like a ring to encircle and slide along DNA (9–11). The γ complex couples ATP to assemble the β ring around DNA (12). The β sliding DNA clamp also binds directly to α, thereby tethering the core polymerase down to DNA for highly processive synthesis (9).

We have a relatively firm understanding of the function of the α (*dnaE*), ε (*dnaQ*), and β (*dnaN*) subunits because they are available in quantity through use of their genes. The only other known holoenzyme subunit gene is *dnaX*, which encodes both τ (71 kDa) and γ (47 kDa); γ is a truncated version of τ due to a translational frameshift (13–15). The carboxyl-terminal sequence of τ provides it with an ability to bind together two molecules of polIII core, presumably for simultaneous synthesis of the leading and lagging strands (16, 17). By itself, γ cannot assemble β onto DNA; for this the δ subunit is also required, and the δ', χ, and ψ subunits stimulate the reaction (12, 18). To gain further insight into the functions of δ, δ', χ, and ψ, their genes must be identified for genetic studies, and the proteins must be produced in quantity for biochemical studies.

In this series of reports the genes encoding δ, δ', χ, and ψ of the γ complex are identified and also the gene encoding θ of polIII core, the last remaining gene of the holoenzyme. We have named these genes *hol*, for holoenzyme (suggested by Dr. Kenneth K. Marians, Sloan-Kettering), and the letters *A–E* in order of descending molecular mass (δ, δ', χ, ψ, and θ, respectively). We have sequenced the genes, cloned them into expression vectors, purified the subunits, and have initiated studies of their structure and function.

We begin the series with the identification of *holA* and *holB* and purification of δ and δ' in quantity. The amino acid sequence of δ' is homologous to the sequence of the γ and τ subunits, revealing a much greater structural redundancy in this multienzyme complex than previously recognized in only the relationship of γ to τ. In the second report, the physical and functional interactions of δ and δ' are characterized alone and with other holoenzyme subunits (19).

EXPERIMENTAL PROCEDURES

E. coli Strains, Bacteriophage, and DNAs

HB101 (pNT203, pSK100) (20) used to purify sufficient δ and δ' for microsequencing was the gift of Dr. Arthur Kornberg (Stanford University). Strain BL21(DE3) plysS and pET3c (21) were gifts of Dr. F. William Studier (Brookhaven National Laboratory). λ15D7 (169) and λE9G1 (236) were gifts of Dr. Yuji Kohara (National Institute of Genetics, Japan). pUC18 and M13mp18 double-strand DNA were from Bethesda Research Laboratories. M13mp18 ssDNA was purified as described (9). All cell growth was in LB medium using ampicillin at 100 μg/ml and chloramphenicol at 30 μg/ml where needed. λ phage was prepared as described (22). Buffer A is 20 mM Tris-HCl (pH 7.5), 20% glycerol, 0.5 mM EDTA, 2 mM DTT. Buffer B is 30 mM Hepes-NaOH (pH 7.2), 10% glycerol, 0.5 mM EDTA, 2 mM DTT.

Microsequencing

The δ and δ' subunits were purified through the ATP-agarose column step (18) from 1.3 kg of HB101(pNT203, pSK100). δ and δ' were separated on a 13% SDS-polyacrylamide gel whereupon δ' resolved into two bands as noted previously (18). The slower and faster migrating δ' bands will be referred to in this report as δ'L

* This work was supported by National Institutes of Health Grant GM38839. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) L04576 (holA) and L04577 (holB).

§ Contributed equally to this work.

‖ To whom correspondence should be sent: Dept. of Microbiology, Cornell University Medical College, 1300 York Ave., New York, NY 10021. Tel.: 212-746-6518; Fax: 212-746-8587.

[1] The abbreviations used are: holoenzyme, DNA polymerase III holoenzyme; kb, kilobase(s); polIII', holoenzyme lacking β; polIII', complex of (α, ε, θ, and τ)$_2$; polIII core, complex of α, ε, θ; γ complex, complex of γ, δ, δ', χ, and ψ; αε polymerase, 1:1 complex of α and ε; ssDNA, single-strand DNA; SSB, *E. coli* ssDNA-binding protein; IPTG, isopropyl-1-thio-β-D-galactopyranoside; DTT, dithiothreitol; bp, base pairs; BSA, bovine serum albumin; HPLC, high performance liquid chromatography.

(large) and δ'_S (small), respectively; δ'_S was approximately two times the abundance of δ'_L by amino acid analysis. δ, δ'_L, and δ'_S were electroblotted onto polyvinylidene difluoride membrane (Whatman) as described (23) for amino-terminal sequencing (50 pmol each) and onto nitrocellulose membrane (Schleicher & Schuell) as described (24) for tryptic analysis (140 pmol of δ, 90 pmol of δ'_L, 180 pmol of δ'_S). Proteins were visualized by Ponceau S stain (Sigma), and their sequences were determined by William S. Lane of Harvard Microchemistry. Sequences of δ were: δ-amino terminus, NH_2-MLRLYPEQLRAQLNEGLRAAYLLLGNDP; tryptic peptides: δ-1, NH_2-AAYLLLGNDPLLLQESQDAVR; δ-2, NH_2-AQENAAWFTALANR; δ-3, NH_2-VEQAVNDAAHFTPFHWVDALLM(G)(K). Sequences of δ'_S were: δ'-amino terminus, NH_2-MRWYPPL(R)(P)DFEKLVA; tryptic peptides: δ'-1, NH_2-EVTEKLNEHAR; δ'-3, NH_2-VVWVTDAALLTDAAANALLK; δ'-4, NH_2-TLEEPPAETWFFLATREP(E)(R)LLAT(L); δ'-5, NH_2-LHYLAPP(P)EQYAVT(W)LSR; δ'-6, NH_2-LSAGSPGAALALFQGDNWQAR. Sequences of tryptic peptides of δ'_L were: δ'-2, NH_2-LGGAK; δ'-7 (same as δ'-3), NH_2-VVWVTDAALLTDAAANALLK. Parentheses indicate uncertain assignments.

Identification of holB

Two synthetic oligonucleotide probes (DNA oligonucleotides, Oligos Etc. Inc.) were designed from the sequence of two of the tryptic peptides and the codon usage of *E. coli* (31) with allowance for a T-G mispair at the wobble position. A synthetic DNA 57-mer based on δ'-4 was 5′-ACTCTGGAAGAACCGCCGGCTGAAACTTGGTTTTTTCTGGCTACTCGTGAACCGGAA-3′ (after identification and sequencing *holB* this probe was incorrect at 11 positions). A DNA 54-mer based on δ'-6 was 5′-GCTGGTTCTCCGGGTGCTGCTCTGGCTCTGTTTCAGGGTGATAACTGGCAGGCT-3′ (the sequence of *holB* showed this probe was incorrect at 9 positions). These probes (100 pmol each) were 5′ end-labeled with 1 μM [γ-^{32}P]ATP (radionucleotides, Du Pont-New England Nuclear) and polynucleotide kinase. *E. coli* genomic DNA (strain C600) was extracted as described (25) and restricted with either *Bam*HI, *Hin*dIII, *Eco*RI, *Eco*RV, *Bgl*I, *Kpn*I, *Pst*I, or *Pvu*II (DNA modification enzymes, New England Biolabs), and then each digest was electrophoresed in a 0.8% native agarose gel followed by Southern analysis (22) using either the 57- or 54-mer as a probe. Initially we used a hybridization temperature of 42 °C and then washed the filters with 2 × SSC and 0.2% SDS at 53 °C. Autoradiography showed a single band in each lane for the 57-mer. The 54-mer showed two bands in each lane, but one band always matched the position of the band probed with the 57-mer.[2]

DNA Sequencing

For the *holA* gene, the 3.2-kb *Kpn*I/*Bgl*II fragment containing *holA* was excised from λ15D7 (169) and directionally ligated into pUC18 to yield pUC-δ. For *holB*, the 2.1-kb *Kpn*I/*Eco*RV fragment containing the *holB* gene was excised from λE9G1 (236) and directionally ligated into pUC18 (*Kpn*I/*Hinc*II) to yield pUC-δ'. Both strands of *holA* and *holB* were sequenced by the chain termination method of Sanger *et al.* (26) using the U. S. Biochemical Corp. Sequenase kit, α-^{35}S-dATP, and synthetic DNA 17-mers.

Overproducing Plasmids

pET-δ—Approximately 1.7 kb of DNA upstream of *holA* was excised from pUC-δ using *Kpn*I (polylinker site) and *Bst*XI (13 bp upstream of the start codon of *holA*) followed by blunt end formation using Klenow polymerase and recircularization of the plasmid using T4 DNA ligase. A 1.5-kb fragment containing *holA* was then excised using *Eco*RI and *Xba*I (these sites are in the pUC18 polylinker on either side of the *holA* insert) followed by directional ligation into M13mp18 to yield M13-δ. An *Nde*I site was generated at the start codon of *holA* by the primer-directed mutagenesis technique of Kunkel *et al.* (27) using a DNA 33-mer (5′-GTACAACCGAATCATATGTTACCCAGCGAGCTC-3′) containing the *Nde*I site (underlined) at the start codon of *holA* and using DNA polymerase III holoenzyme and SSB to replicate the circle completely without strand displacement as described by O'Donnell and Kornberg (28). An *Nde*I fragment (2.1 kb) containing *holA* was excised from the *Nde*I-mutated M13-δ (M13-δ_{Nde}) and ligated into pET3c, linearized using *Nde*I, to yield pET-δ. The orientation of *holA* in pET-δ was determined by sequencing.

pET-δ'—A 2.1-kb *Kpn*I/*Hin*dIII fragment containing *holB* was excised from pUC-δ' and directionally ligated into M13mp18 to yield M13-δ'. An *Nde*I site was generated at the start codon of *holB* using a DNA 33-mer (5′-GGTGAAGGAGTTGGACATATGAGATGGTATCCA-3′) containing the *Nde*I site (underlined) at the start codon of *holB* as described above for *holA*. An *Nde*I fragment (1,160 bp) containing *holB* was excised from M13-δ'_{Nde} and ligated into pET3c to yield pET-δ' as described above.

Replication Assay for δ

The δ replication assay contained 72 ng of M13mp18 ssDNA (0.03 pmol as circles) uniquely primed with a DNA 30-mer (9), 980 ng of SSB (13.6 pmol as tetramer), 22 ng of β (0.29 pmol as dimer), 200 ng of γ (2.1 pmol as dimer), 55 ng of $\alpha\epsilon$ complex (0.35 pmol) in a final volume (after the addition of proteins) of 25 μl of replication assay buffer (20 mM Tris-HCl (pH 7.5); 8 mM $MgCl_2$; 5 mM DTT; 4% glycerol; 40 μg/ml BSA; 0.5 mM ATP; a 60 μM concentration each of dCTP, dGTP, and dATP; and 20 μM [α-^{32}P]dTTP). Proteins used in the assay were purified as described (17). 1–5 ng of δ (or column fraction) was added to the assay on ice, shifted to 37 °C for 5 min, and then quenched and quantitated using DE81 paper as described (29). When needed, proteins were diluted in buffer A containing 60 mM NaCl and 50 μg/ml BSA.

Replication Assay for δ'

The δ' replication assay contained 108 ng of M13mp18 ssDNA (0.05 pmol as circles) primed with a DNA 30-mer (9), 1.5 μg of SSB (21 pmol as tetramer), 30 ng of β (0.39 pmol as dimer), 22.5 ng of $\alpha\epsilon$ complex (0.14 pmol), 20 ng of γ (0.21 pmol as dimer), and 2 ng of δ ((0.05 pmol as monomer) in a final volume of 25 μl of assay buffer lacking dATP and dTTP. 1–5 ng of δ' (or column fraction) was added to the assay on ice then shifted to 37 °C for 8 min to allow assembly of the processive polymerase. DNA synthesis was initiated upon rapid addition of 60 μM dATP, 20 μM [α-^{32}P]TTP and then quenched after 20 s and quantitated using DE81 paper as described (29). When needed, proteins were diluted in buffer A containing 50 μg/ml BSA.

RESULTS

Identification of holA—The sequence of the amino-terminal 28 amino acids of δ and three internal tryptic peptides were determined. One of the tryptic peptides (21 amino acids) overlapped 10 amino acids of the amino-terminal sequence. A search of the translated GenBank (30) revealed an exact match to the 21-amino acid tryptic peptide which overlapped the amino-terminal sequence. The matching sequence occurred just downstream of the *rlpB* gene at 15.2 min (688 kb starting from *thrA*) of the *E. coli* chromosome (32). The match to the amino-terminal sequence of δ was imperfect because of a few errors in the published sequence of this region.[3] The published sequence information downstream of *rlpB* accounted for approximately 22% of the δ gene and did not encode the other two tryptic fragments.

A 3.2-kb *Kpn*I/*Bgl*II fragment containing *holA* was excised from Kohara phage λ15D7 (169), cloned into pUC18, and the *holA* gene was sequenced (Fig. 1). The *holA* sequence encodes the correct amino terminus of δ and all of the tryptic peptides in the same reading frame (*underlined* in Fig. 1). Overall, *holA* encodes a 343-amino acid protein (pI = 6.99) of 38,704 Da, consistent with the mobility of δ in SDS-polyacrylamide gels (20). The termination codon of the *rlpB* gene overlaps the initiating ATG of *holA*. It is not known whether *holA* is in an operon with *rlpB* or a downstream gene. One nucleotide beyond the *holA* stop codon is a possible ATG initiation codon for an open reading frame throughout the rest of the known

[2] We cloned and sequenced both genes to which the 54-mer probe hybridized. One was *holB*, the other contained 15 nucleotides in a row which exactly matched the 54-mer, but there was no amino acid homology to δ' in the sequence surrounding these 5 amino acids.

[3] The previously published sequence (32) encoding the first 54 nucleotides of *holA* (*i.e.* the amino-terminal 18 amino acids of δ) differed by 11 nucleotides from the sequence presented here and was a particularly difficult stretch of DNA to sequence.

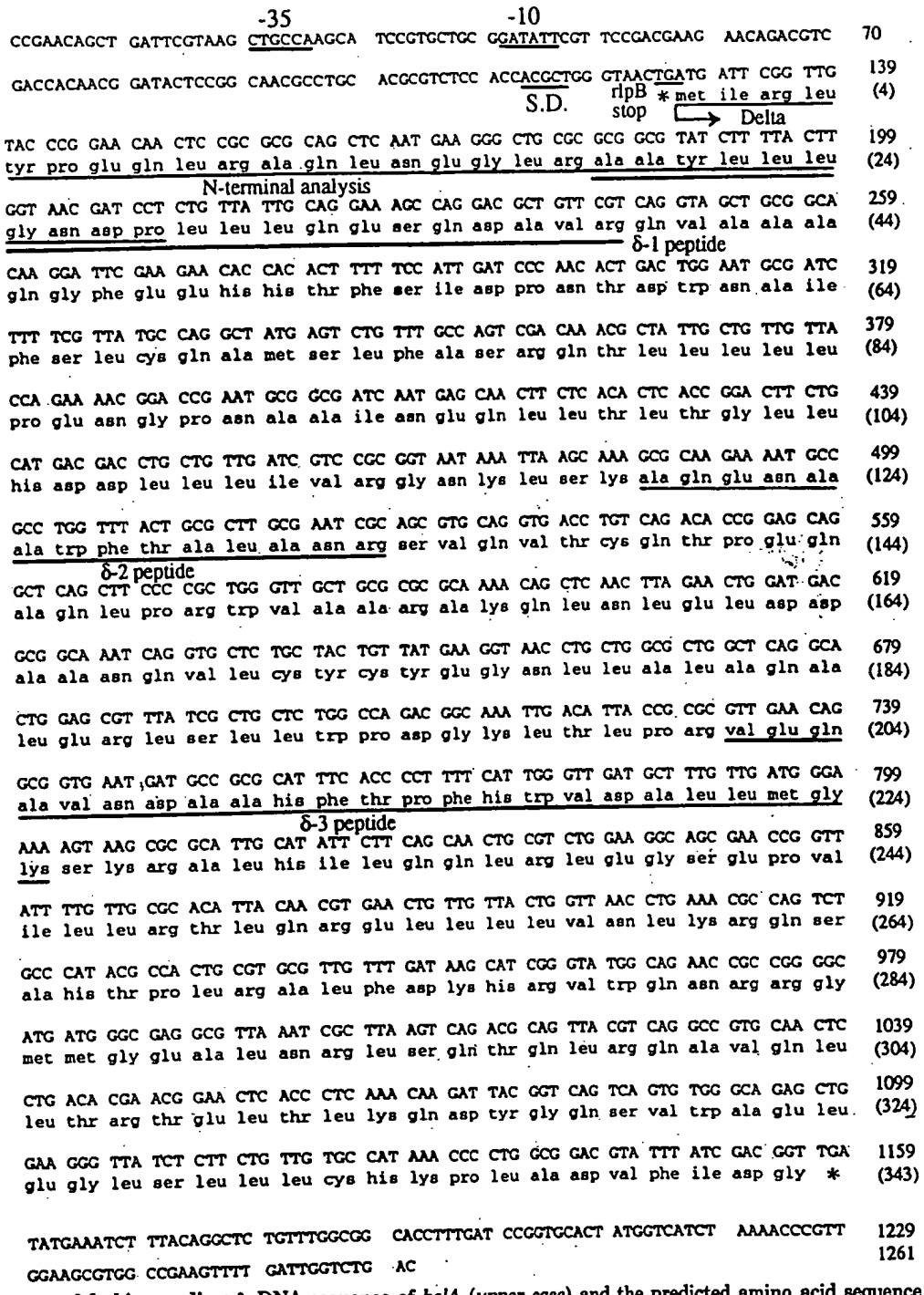


FIG. 1. Sequence of *holA* encoding δ. DNA sequence of *holA* (*upper case*) and the predicted amino acid sequence of δ (*lower case*) predicts a 343-amino acid protein of 38,704 Da. The amino-terminal and tryptic peptide sequences obtained from the naturally purified δ subunit are *underlined*. The stop codons of *rlpB* and *holA* are marked with *asterisks*. The putative RNA polymerase promoter signals (−35, −10) and Shine-Dalgarno (*S.D.*) sequence are *underlined*. Numbering of the nucleotide sequence is presented to the *right*. Numbering of the amino acids of δ is shown in *parentheses* to the *right*.

sequence (101 bp), suggesting a possible gene immediately following *holA*. The nearest possible initiation signals for transcription and translation of the *holA* gene are *underlined* in Fig. 1; the match to their consensus sequences is not strong, suggesting a low utilization efficiency. Inefficient transcription and/or translation may be expected for a gene encoding a subunit of the holoenzyme present at only 10–20 copies/cell (33). The *holA* gene uses several rare codons (CCC(Pro), ACA(Thr), GGA(Gly), AGT(Ser), AAT(Asn), TTA, TTG, CTC (all Leu)) two to five times more frequently than average (31), which may decrease translation efficiency.

The amino-terminal region of δ (amino acids 17–38) contains a possible leucine zipper ($LeuX_6LeuX_6LeuX_6Val$) although a proline between the second and third leucines may disrupt the required α-helical structure. Of the 33 arginine and lysine residues in δ, 16 (50%) are within amino acids 225–307. This same region contains only 5 (14%) of the 35 glutamic and aspartic acid residues. Whether this concentration of basic residues is significant to function is unknown. There are no zinc finger or helix-turn-helix motifs.

ATP binding to δ has been detected previously by a UV cross-linking study of the holoenzyme (34). The sequence of

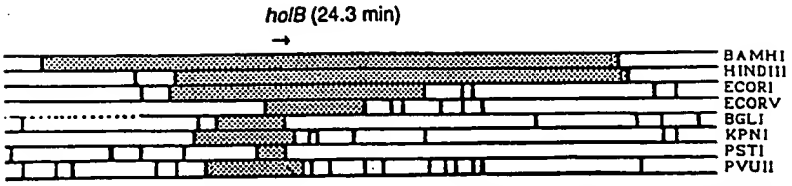


FIG. 2. **Position of *holB* on the Kohara map.** Section of the Kohara restriction map in the vicinity of 24 min. Restriction enzymes are indicated to the *right*. *Shaded* fragments indicate those detected by Southern analysis. The position of overlap places *holB* at 24.3 min. The direction of *holB* is based on the relative positions of the *Kpn*I site to the *Eco*RV site in the Kohara map and the known direction of *holB* within this fragment by sequence analysis. The correspondence between the observed fragment sizes with those on the map (observed *versus* map) were (in kb): *Bam*HI, >25 *versus* 38; *Hin*dIII, >20 *versus* 30; *Eco*RI, >15 *versus* 16.2; *Eco*RV, 7.0 *versus* 6.8; *Bgl*I, 4.2 *versus* 4.2; *Kpn*I, 6.6 *versus* 6.4; *Pst*I, 1.7 *versus* 1.9; *Pvu*II, 6.2 *versus* 6.2.

```
AAGAATCTTT  CGATTTCTTT  AATCGCACCC GCGCCCGCTA TCTGGAACTG GCAGCACAAG ATAAAAGCAT  70

TCATACCATT  GATGCCACCC  AGCCGCTGGA GGCCGTGATG GATGCAATCC GCACTACCGT GACCCACTGG  140

GTGAAGGAGT  TGGACGCATG AGA TGG TAT CCA TGG TTA CGA CCT GAT TTC GAA AAA CTG GTA  202
S.D.              met arg trp tyr pro trp leu arg pro asp phe glu lys leu val  (15)
                  ↳ Delta prime          N-terminal analysis
GCC AGC TAT CAG GCC GGA AGA GGT CAC CAT GCG CTA CTC ATT CAG GCG TTA CCG GGC ATG  262
ala ser tyr gln ala gly arg gly his his ala leu leu ile gln ala leu pro gly met  (35)

GGC GAT GAT GCT TTA ATC TAC GCC CTG AGC CGT TAT TTA CTC TGC CAA CAA CCG CAG GGC  322
gly asp asp ala leu ile tyr ala leu ser arg tyr leu leu cys gln gln pro gln gly  (55)

CAC AAA AGT TGC GGT CAC TGT CGT GGA TGT CAG TTG ATG CAG GCT GGC ACG CAT CCC GAT  382
his lys ser cys gly his cys arg gly cys gln leu met gln ala gly thr his pro asp  (75)

TAC TAC ACC CTG GCT CCC GAA AAA GGA AAA AAT ACG CTG GGC GTT GAT GCG GTA CGT GAG  442
tyr tyr thr leu ala pro glu lys gly lys asn thr leu gly val asp ala val arg glu  (95)

GTC ACC GAA AAG CTG AAT GAG CAC GCA CGC TTA GGT GGT GCG AAA GTC GTT TGG GTA ACC  502
val thr glu lys leu asn glu his ala arg leu gly gly ala lys val val trp val thr  (115)
                δ'-1 peptide                δ'-2 peptide
GAT GCT GCC TTA CTA ACC GAC GCC GCG GCT AAC GCA TTG CTG AAA ACG CTT GAA GAG CCA  562
asp ala ala leu leu thr asp ala ala ala asn ala leu leu lys thr leu glu glu pro  (135)
                δ'-3 peptide
CCA GCA GAA ACT TGG TTT TTC CTG GCT ACC CGC GAG CCT GAA CGT TTA CTG GCA ACA TTA  622
pro ala glu thr trp phe phe leu ala thr arg glu pro glu arg leu leu ala thr leu  (155)
                                δ'-4 peptide
CGT AGT CGT TGT CGG TTA CAT TAC CTT GCG CCG CCG CCG GAA CAG TAC GCC GTG ACC TGG  682
arg ser arg cys arg leu his tyr leu ala pro pro pro glu gln tyr ala val thr trp  (175)
                                δ'-5 peptide
CTT TCA CGC GAA GTG ACA ATG TCA CAG GAT GCA TTA CTT GCC GCA TTG CGC TTA AGC GCC  742
leu ser arg glu val thr met ser gln asp ala leu leu ala ala leu arg leu ser ala  (195)

GGT TCG CCT GGC GCG GCA CTG GCG TTG TTT CAG GGA GAT AAC TGG CAG GCT CGT GAA ACA  802
gly ser pro gly ala ala leu ala leu phe gln gly asp asn trp gln ala arg glu thr  (215)
                                δ'-6 peptide
TTG TGT CAG GCG TTG GCA TAT AGC GTG CCA TCG GGC GAC TGG TAT TCG CTG CTA GCG GCC  862
leu cys gln ala leu ala tyr ser val pro ser gly asp trp tyr ser leu leu ala ala  (235)

CTT AAT CAT GAA CAA GCT CCG GCG CGT TTA CAC TGG CTG GCA ACG TTG CTG ATG GAT GCG  922
leu asn his glu gln ala pro ala arg leu his trp leu ala thr leu leu met asp ala  (255)

CTA AAA CGC CAT CAT GGT GCT GCG CAG GTG ACC AAT GTT GAT GTG CCG GGC CTG GTC GCC  982
leu lys arg his his gly ala ala gln val thr asn val asp val pro gly leu val ala  (275)

GAA CTG GCA AAC CAT CTT TCT CCC TCG CGC CTG CAG GCT ATA CTG GGG GAT GTT TGC CAC  1042
glu leu ala asn his leu ser pro ser arg leu gln ala ile leu gly asp val cys his  (295)

ATT CGT GAA CAG TTA ATG TCT GTT ACA GGC ATC AAC CGC GAG CTT CTC ATC ACC GAT CTT  1102
ile arg glu gln leu met ser val thr gly ile asn arg glu leu leu ile thr asp leu  (315)

TTA CTG CGT ATT GAG CAT TAC CTG CAA CCG GGC GTT GTG CTA CCG GTT CCT CAT CTT TAA  1162
leu leu arg ile glu his tyr leu gln pro gly val val leu pro val pro his leu  *  (334)

 GAGAGACATC ATGTTTTTAG TCGACTCACA  CTGCCATCTC GATGGTCTGG ATTATGAATC  TTTGCATAAG  1232

 GACGTGGATG ACGTTCTGGC GAAAGCCGCC  GCACGCGATG TGAAATTTTG TCTGGCAGTC  GCCACAACAT  1302
```

FIG. 3. **Sequence of *holB* encoding δ'.** DNA sequence of *holB* (*upper case*) and the predicted amino acid sequence of δ' (*lower case*) predicts a 334-amino acid protein of 36,937 Da. The amino-terminal and tryptic peptide sequences obtained from the naturally purified δ' are *underlined*. The stop codon is marked with an *asterisk*. The putative translational signal (Shine-Dalgarno (*S.D.*)) is *underlined*. Numbering of the nucleotide sequence and amino acid sequence (in *parentheses*) is presented to the *right*.

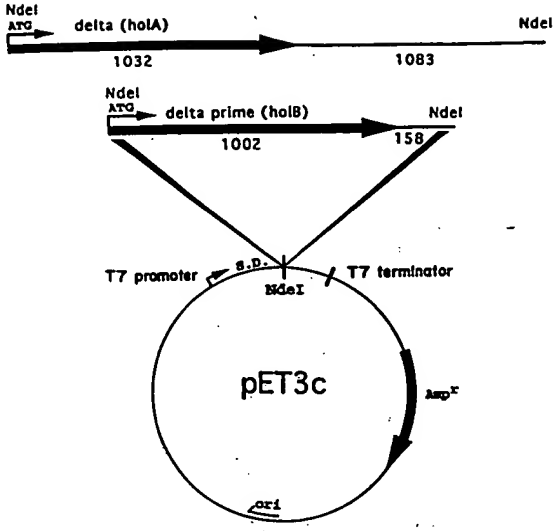


FIG. 4. **The δ and δ′ expression vectors.** Construction of pET-δ and pET-δ′ is described under "Experimental Procedures." The *holA* insert in pET-δ and *holB* insert in pET-δ′ are shown above the pET3c expression vector. The initiating ATG of these genes is positioned downstream of the Shine-Dalgarno sequence (*S.D.*) and a T7 promoter. Downstream of *holA* are 492 bp of *E. coli* DNA and 591 bp of M13mp18 DNA. The *holB* insert contains 158 bp of *E. coli* DNA downstream of *holB* to an *Nde*I site. The T7 RNA polymerase termination sequence is downstream of the insert.

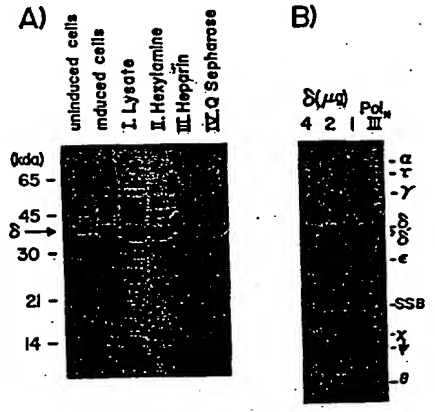


FIG. 5. **Purification of δ.** *Panel A*, Coomassie Blue-stained 13% SDS-polyacrylamide gel of column pools. *First lane*, total cells, uninduced. *Second lane*, total cells, induced. *Third lane*, cell lysate supernatant (70 μg). *Fourth lane*, hexylamine-Sepharose pool (15 μg). *Fifth lane*, heparin-agarose pool (7 μg). *Sixth lane*, Q-Sepharose pool (7 μg). Positions of weight markers are to the *left*. *Panel B*, comparison of the mobility of cloned δ (*first three lanes*, 4, 2, and 1 μg, respectively) with δ in polIII* (*fourth lane*, 4 μg). Subunits of polIII* and the SSB which copurified with it are identified to the *right*.

holA shows a near match to the ATP binding site consensus motif (*i.e.* AX_5GKS for δ at residues 219–225 compared with the consensus sequences G/AX_4GKT/S, G/AXGKT/S, or G/AX_2GXGKT/S, Ref. 35). Whether δ binds ATP specifically at this site remains to be determined.

Identification of holB—The naturally purified δ′ appears in a 13% SDS-polyacrylamide gel as two closely spaced bands (18). The slower migrating species (δ'_L) is approximately one-half the abundance of the faster migrating species (δ′s). Both δ'_L and δ′s are probably encoded by the same gene as their HPLC profiles upon digestion with trypsin were very similar (data not shown). In support of this, peptides from δ′s and δ'_L which had the same retention time on HPLC analysis also had identical amino acid sequences (peptide δ′-7 from δ′s and δ′-3 from δ'_L were identical). The amino terminus and five tryptic peptides of δ′s and two tryptic peptides of δ'_L were sequenced. A search of the translated GenBank (30) revealed no match to any of the δ′ sequences.

The *holB* gene was identified using two oligonucleotide probes in a Southern analysis of *E. coli* genomic DNA digested with the eight Kohara restriction map enzymes. Imposing the restraint that the eight fragments from the Southern analysis must overlap at the *holB* gene, the Kohara map of the *E. coli* chromosome (36) was searched, and only one position of overlap of fragments of the observed size was present on the map. This position was located at 24.3 min (1,174 kb starting from *thrA*) (Fig. 2). A restriction fragment containing the overlapping region was subcloned from λE9G1 (236) and sequenced (Fig. 3). The open reading frame encodes the amino-terminal sequence and all six tryptic peptides obtained from δ'_L and δ′s (*underlined* in Fig. 3).

The *holB* gene encodes a 334-amino acid protein of 36,937

TABLE I

Purification of δ

BL21(DE3) pET-δ cells were grown at 37 °C in 12 liters of LB containing 1.2 g of ampicillin. Upon growth to an A_{600} of 1.5 the temperature was lowered to 25 °C, and IPTG was added to 0.4 mM. After a further 3 h the cells (50 g) were collected by centrifugation. Cells were lysed using lysozyme as described (42), and the debris was removed by centrifugation. The purification steps that follow were performed at 4 °C. The assay for δ is described under "Experimental Procedures." The clarified cell lysate (Fraction I, 300 ml) was diluted 2-fold with buffer A to a conductivity equal to 112 mM NaCl then loaded onto a 60-ml hexylamine-Sepharose column (Pharmacia LKB Biotechnology Inc.) equilibrated with buffer A + 0.1 M NaCl. The hexylamine column was washed with 60 ml of buffer A + 0.1 M NaCl then eluted over a period of 14 h using a 600-ml linear gradient of 0.1–0.5 M NaCl in buffer A. Eighty fractions were collected. Fractions 16–34 (Fraction II, 125 ml) were dialyzed against 2 liters of buffer A + 90 mM NaCl overnight then diluted 2-fold with buffer A to yield a conductivity equal to 65 mM NaCl just prior to loading onto a 60-ml column of heparin-Sepharose (Pharmacia) equilibrated in buffer A + 50 mM NaCl. The heparin-Sepharose column was washed with 120 ml of buffer A + 50 mM NaCl and then eluted over a period of 14 h using a 600-ml linear gradient of 0.05–0.5 M NaCl in buffer A. Eighty fractions were collected. Fractions 24–34 (Fraction III) were pooled and diluted 3-fold (final volume, 250 ml) with buffer A to a conductivity equal to 85 mM NaCl just prior to loading onto a 50-ml Hi Load 26/10 Q-Sepharose fast flow fast protein liquid chromatography column (Pharmacia). The Q-Sepharose column was washed with 150 ml of buffer A + 50 mM NaCl and then eluted over a period of 10 h using a 600-ml linear gradient of 0.05–0.5 M NaCl in buffer A. Eighty fractions were collected. Fractions 28–36 were pooled (Fraction IV, 74 ml, 1.9 mg/ml) and passed over a 1-ml ATP-Sepharose column (Pharmacia, type II (N-6 linked)) to remove any possible γ complex contaminant and then dialyzed *versus* two changes of 2 liters each of buffer A containing 0.1 M NaCl (the DTT was omitted for the purpose of determining protein concentration spectrophotometrically) before storing at −70 °C. Protein concentration was determined by the Bradford method (43) using BSA as a standard except at the last step in which the concentration was determined by absorbance using $\epsilon_{280} = 46{,}137\ M^{-1}\ cm^{-1}$.

Fraction	Total protein	Total units[a]	Specific activity	Fold purification	Yield
	mg		*units/mg*		%
I. Lysate[b,c]	2,070	5.4×10^9	2.6×10^6	1.0	100
II. Hexylamine	446	2.5×10^9	5.6×10^6	2.2	46
III. Heparin	197	2.0×10^9	10.2×10^6	3.9	37
IV. Q-Sepharose[d]	141	1.5×10^9	10.6×10^6	4.1	28

[a] One unit is defined as 1 pmol of nucleotide incorporated per min.

[b] Lysate of BL21(DE3) cells harboring the pET3c vector yielded a specific activity of 10^4 units/mg.

[c] Omission of γ from the assay of the lysate resulted in a 200-fold reduction of specific activity (1.2×10^4 units/mg).

[d] Omission of γ from the assay using pure δ gave no detectable synthesis.

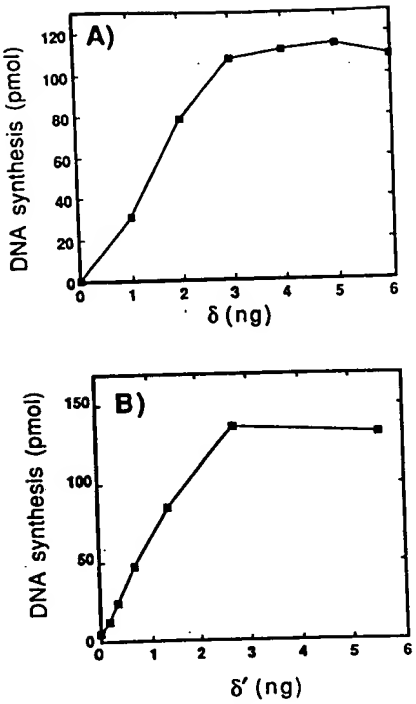


FIG. 6. **Replication activity of δ and δ′.** *Panel A*, the pure δ subunit was titrated into the replication assay which is described under "Experimental Procedures." *Panel B*, the pure δ′ subunit was titrated into the replication assay which is described under "Experimental Procedures."

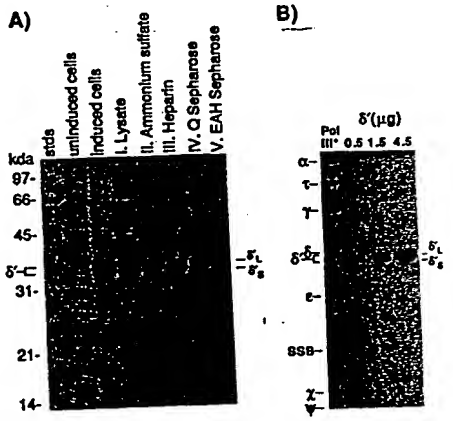


FIG. 7. **Purification of δ′.** *Panel A*, Coomassie Blue-stained 13% SDS-polyacrylamide gel of column pools. *First lane*, molecular mass standards. *Second lane*, total cells, uninduced. *Third lane*, total cells, induced. *Fourth lane*, cell lysate supernatant (41 μg). *Fifth lane*, ammonium sulfate pool (8 μg). *Sixth lane*, heparin-agarose pool (6 μg). *Seventh lane*, Q-Sepharose pool (6 μg). *Eighth lane*, EAH-Sepharose pool (6 μg). Molecular mass of the markers is to the *left*, and positions of δ'_L and δ'_S are indicated to the *right*. *Panel B*, comparison of the mobility of cloned δ′ (*second through fourth lanes*, 0.5, 1.5, and 4.5 μg, respectively) with δ′ within polIII* (*first lane*, 4 μg). Subunits of polIII* (θ ran off) and SSB which copurified with it are identified on the *left side* of the gel, and δ'_L and δ'_S are identified to the *right*.

Da (pI = 7.04), consistent with the mobility of δ′ in an SDS-polyacrylamide gel (18, 20). Upstream of *holB* is a putative Shine-Dalgarno sequence (Fig. 3). It is not known whether *holB* is in an operon. The possibility of a gene directly upstream of *holB* is indicated by an open reading frame (+1 relative to *holB*) throughout the 158-bp upstream sequence which terminates with a TGA stop codon that overlaps the initiating ATG of *holB*. In addition, there is an ATG 10 bp downstream of *holB* without an in frame stop codon over the remaining 130 bp of sequence suggesting a possible downstream gene. There is no obvious promoter for *holB*, consistent with the possibility that the *holB* gene is in an operon. Alternatively, the promoter may poorly match the consensus, as a low level of transcription may be expected for a subunit of a holoenzyme which is present at only low levels. The *holB* gene uses several rare codons (TTA (Leu), ACA (Thr), GGA (Gly), AGC, TCG (Ser)) two to four times more frequently than average (31) which may decrease translation efficiency.

The *holB* gene contains a helix-turn-helix motif (Ala/

TABLE II

Purification of δ′

300 liters of BL21(DE3) plysS, pET-δ′ cells were grown at 37 °C in LB supplemented with 5 mg/ml glucose, 10 μg/ml thiamine, 50 μg/ml thymine containing 100 μg/ml ampicillin, and 25 μg/ml chloramphenicol. Upon growth to an A_{600} of 0.5, IPTG was added to 0.2 mM. After further growth for 2 h the cells (940 g) were collected by centrifugation, resuspended in an equal weight of 50 mM Tris-HCl (pH 7.5), 10% sucrose, and stored at −70 °C. 100 g of cells (30 liters of cell culture) were thawed whereupon they lysed (because of lysozyme produced by plysS), and cell debris was removed as described (42) to yield the cell lysate (Fraction I, 4.41 g in 325 ml). The purification steps that follow were performed at 4 °C. The assay for δ′ is described under "Experimental Procedures." Ammonium sulfate (0.21 g/ml) was added to step I and stirred for 90 min. The pellet contained δ′ (Fraction II, 1.58 g) and was redissolved in 660 ml of buffer B and dialyzed against two changes of 2 liters each of buffer B to a conductivity equal to 40 mM NaCl. The Fraction II was loaded onto a 300-ml heparin-agarose column (Bio-Rad) equilibrated with buffer B. The heparin column was washed with 450 ml of buffer B + 20 mM NaCl and then eluted over a period of 14 h using a 2.5-liter linear gradient of 20–300 mM NaCl in buffer B. One hundred fractions were collected. Fractions 36–53 were pooled (Fraction III, 550 ml) and dialyzed twice against 2 liters of buffer A to a conductivity equal to 60 mM NaCl. The Fraction III was loaded onto a 100-ml Q-Sepharose column (Pharmacia) equilibrated with buffer A. The Q-Sepharose column was washed with 150 ml of buffer A + 20 mM NaCl and then eluted over a period of 12 h using a 1.2-liter linear gradient of 20–300 mM NaCl in buffer A. Eighty fractions were collected. Fractions 34–56 were pooled (Fraction IV, 370 ml) and dialyzed twice against 2 liters each of buffer A to a conductivity equal to 60 mM NaCl just prior to loading onto a 60-ml EAH-Sepharose column (Pharmacia's replacement for hexylamine-Sepharose) equilibrated with buffer A. The EAH-Sepharose column was washed with 60 ml of buffer A + 40 mM NaCl and then eluted over a period of 10 h with a 720-ml linear gradient of 40–500 mM NaCl in buffer A. Eighty fractions were collected. Fractions 18–30 (Fraction V, 130 ml), which contained homogeneous δ′, were pooled and dialyzed against 2 liters of buffer A (lacking DTT to allow an absorbance measurement, see below) to a conductivity of 40 mM NaCl. The Fraction V was passed over a 5-ml ATP-agarose column (Pharmacia, type II, N-6 linked) to remove any γ complex contaminant followed by the addition of DTT to 2 mM and then was aliquoted and stored at −70 °C. Protein concentration was determined by the method of Bradford (43) using BSA as a standard except at the last step in which concentration was determined by absorbance.

Fraction	Total protein	Total units[a]	Specific activity	Fold purification	Yield
	mg		*units/mg*		%
I. Lysate[b,c]	4,414	3.0×10^{1}	7×10^{6}	1.0	100
II. Ammonium sulfate	1,584	2.5×10^{10}	16×10^{6}	2.3	83
III. Heparin	990	2.6×10^{10}	26×10^{6}	3.7	87
IV. Q-Sepharose	781	2.6×10^{10}	33×10^{6}	4.7	87
V. EAH-Sepharose[d]	732	2.5×10^{10}	34×10^{6}	4.9	83

[a] One unit is defined as 1 pmol nucleotide incorporated in 20 seconds.

[b] Lysate of BL21(DE3) plysS cells harboring the pET3c vector yielded a specific activity of 1252 units/mg.

[c] Omission of γ and δ from the assay of the lysate resulted in a 7,650-fold reduction of specific activity (915 units/mg).

[d] Using pure δ′, omission of γ from the assay gave no detectable synthesis under the conditions of the assay.

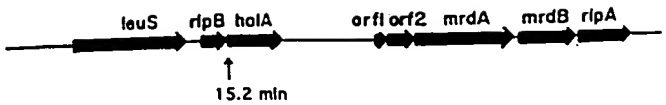


FIG. 8. Genes in the vicinity of *holA* on the *E. coli* chromosome. The *holA* gene is located between the *rlpB* and *mrdA* genes at 15.2 min on the *E. coli* chromosome. *leuS*, leucyl-tRNA synthase; *rlpB*, rare lipoprotein B; *holA*, δ subunit of DNA polymerase III holoenzyme; *orf1*, open reading frame for a 7.7-kDa protein (32); *orf2*, open reading frame for a 17.7-kDa protein (32)); *mrdA* (also *pbpA*), penicillin-binding protein 2; *mrdB* (also *rodA*) *rodA* protein; *rlpA*, rare lipoprotein A.

$GlyX_3GlyX_6IIe/Val$) at $Ala_{80}Gly_{84}Val_{90}$ although ability of δ′ to bind DNA has yet to be examined. There is also a possible leucine zipper ($Leu_7X_6Leu_{14}X_6Gly_{21}X_6Leu_{28}$) in the amino terminus although Gly interrupts the Leu pattern. The *holB* sequence does not contain an ATP binding site motif or a zinc finger.

Purification of δ—The *holA* gene was cloned into M13mp18 followed by site-directed mutagenesis to create an *Nde*I site at the initiating methionine. The *holA* gene was then subcloned into the *Nde*I site of the pET3c expression vector (21) which places *holA* under control of a strong T7 RNA polymerase promotor (Fig. 4). pET-δ was transformed into BL21(DE3) cells which harbor a λ lysogen containing the T7 RNA polymerase gene controlled by the lacUV5 promoter (21). Upon induction of T7 RNA polymerase with IPTG, δ was expressed to 27% total cell protein (compare Fig. 5*A*, *first lane* (uninduced) and *second lane* (induced)). We purified 141 mg of δ (Table I) starting from 12 liters of induced cells (soluble fraction, Fig. 5*A*, *third lane*) upon column fractionation using hexylamine-Sepharose (Fig. 5*A*, *fourth lane*), heparin agarose (*fifth lane*), and Q-Sepharose (*sixth lane*). δ tended to precipitate upon standing in low salt (<70 mM), especially during dialysis. Therefore, low salt was avoided except for short periods of time, and column fractions containing δ were sometimes diluted for the next column rather than dialyzed overnight. Cloned δ comigrated with δ within polIII* (holoenzyme lacking β) (Fig. 5*B*). The δ subunit was assayed as described in an earlier study (18), which demonstrated its requirement to reconstitute the rapid and processive polymerase from the β and γ subunits and αε polymerase (Fig. 6*A*). The ϵ_{280} value calculated from the *holA* sequence is 46,230 M^{-1} cm^{-1}.[4] The measured absorbance of δ in 6 M guanidine hydrochloride is only 0.2% higher than in its native state. Hence, the ϵ_{280} of native δ is 46,137 M^{-1} cm^{-1}.

Purification of δ′—*holB* was cloned into pET3c (Fig. 4) and transformed into BL21 (DE3) plysS. Upon induction of resident T7 RNA polymerase with IPTG, δ′ was expressed to 50% of total cell protein (compare Fig. 7*A*, *second lane* (uninduced cells) and *third lane* (induced cells)). Upon cell lysis, the majority of δ′ was soluble (Fig. 7*A*, *fourth lane*). We purified over 700 mg of δ′ (Table II) starting from 30 liters of cell culture. The cell lysate was fractionated first by ammonium sulfate precipitation (*fifth lane*) and then by successive column fractionation steps using heparin-agarose (*sixth lane*), Q-Sepharose (*seventh lane*), and EAH-Sepharose (*eighth lane*). The δ′ subunit was assayed by its ability to stimulate reconstitution of the highly processive polymerase with the γ, δ, and β subunits and αε polymerase (12). In the absence of δ′, this assay requires a large excess of γ subunit (described in the accompanying report, Ref. 19). Using a low concentration of γ, the δ′ stimulates the assay 30-fold (Fig. 6*B*). The ϵ_{280} value of δ′ calculated from the *holB* sequence is 59,600 M^{-1} cm^{-1}, only 0.9% lower than in the presence of 6 M guanidine hydrochloride for a native ϵ_{280} value of 60,136 M^{-1} cm^{-1}.[4]

The overproduced δ′ consists of a doublet just as observed previously for naturally purified δ′ (18). Fig. 7*B* shows that the two polypeptides of cloned δ′ comigrate with those of δ′ in polIII*. As with naturally purified δ′, the abundance of δ'_L is approximately 50% the abundance of δ'_S. A trivial explanation for the two polypeptides is that (δ'_S) is a proteolytic product of δ'_L. However, electrospray mass spectrometry showed the major species, δ'_S, had a mass of 36,930 Da, which is the mass predicted from the entire *holB* gene sequence, indicating that δ'_S is not the result of proteolytic degradation.[5] The nature of δ'_L is presently under investigation.

DISCUSSION

Chromosomal Context of holA—It is interesting to note that *holA* is in an area of the chromosome containing several membrane protein genes (Fig. 8). They are all transcribed in the same direction. The *mrdA* and *mrdB* genes encode proteins responsible for the rod shape of *E. coli*, and the *rlpA*, *rlpB* genes encode rare lipoproteins speculated to be important to cell duplication (32). The position of *holA* within a cluster of membrane proteins may be coincidental or may help coordinate duplication of the cell and the chromosome.

The δ′ Doublet—δ′ in the γ complex and the cloned δ′ appear as a doublet in a 13% SDS-polyacrylamide gel. Electrospray mass spectrometry of the faster migrating species, δ'_S, showed it was the size of the full-length gene and not derived by proteolysis, suggesting that δ'_L, which migrates slower than δ'_S, may be a modified form of increased size. Possible modifications include mRNA splicing, use of an upstream ATG, read-through of the stop codon, translational frame shifting, and covalent modification. Amino-terminal sequence analysis of the cloned δ'_L and δ'_S subunits showed them to have identical amino termini, proving that δ'_L is not derived from an alternate upstream ATG start site (data not shown). Treatment with calf intestinal and bacterial alkaline phosphatases did not effect the mobility of either δ'_S or δ'_L, suggesting that serine and threonine phosphorylation are not involved (not shown) although phosphorylation of other residues or other types of covalent modification remains a possibility. Translational read-through would produce a protein containing 19 additional residues for an increase of 2130 Da. A −1 translational frame shift would produce a protein containing an additional 7 amino acids before encountering a stop codon in the −1 frame. We detect no obvious frame shifting signals in *holB* (*i.e.* a stretch of A residues followed by a hairpin; or a *XXX YYY Z* site).

δ′ Is Homologous to γ and τ—A homology search of the translated GenBank showed that the most homologous protein to δ′ was another *E. coli* protein, the γ/τ subunit(s) of DNA polymerase III holoenzyme. There is 27% identity and 44% similarity including conservative substitutions over the entire lengths of δ′ and γ/τ (Fig. 9). One particular region in δ′ of 50 amino acids (amino acids 110–159) is strikingly similar to γ/τ (amino acids 121–170) having 49% identity.

The extent of sequence homology between δ′ and γ/τ subunits is above the level required to speculate that they have similar three-dimensional structures. What function does such structural redundancy within one multiprotein ma-

[4] The ϵ_{280} value calculated from the amino acid composition of a protein is within 5% of the ϵ_{280} value in the native state using the equation: $\epsilon_{280} = Trp_m$ (5690 M^{-1} cm^{-1}) + Tyr_n (1280 M^{-1} cm^{-1}) where m and n are the number of Trp and Tyr residues in δ, respectively (39).

[5] Electrospray mass spectrometry of δ'_S was performed by Dr. William S. Lane of Harvard Microchemistry. The δ'_S was separated from δ'_L by reverse phase HPLC prior to analysis.

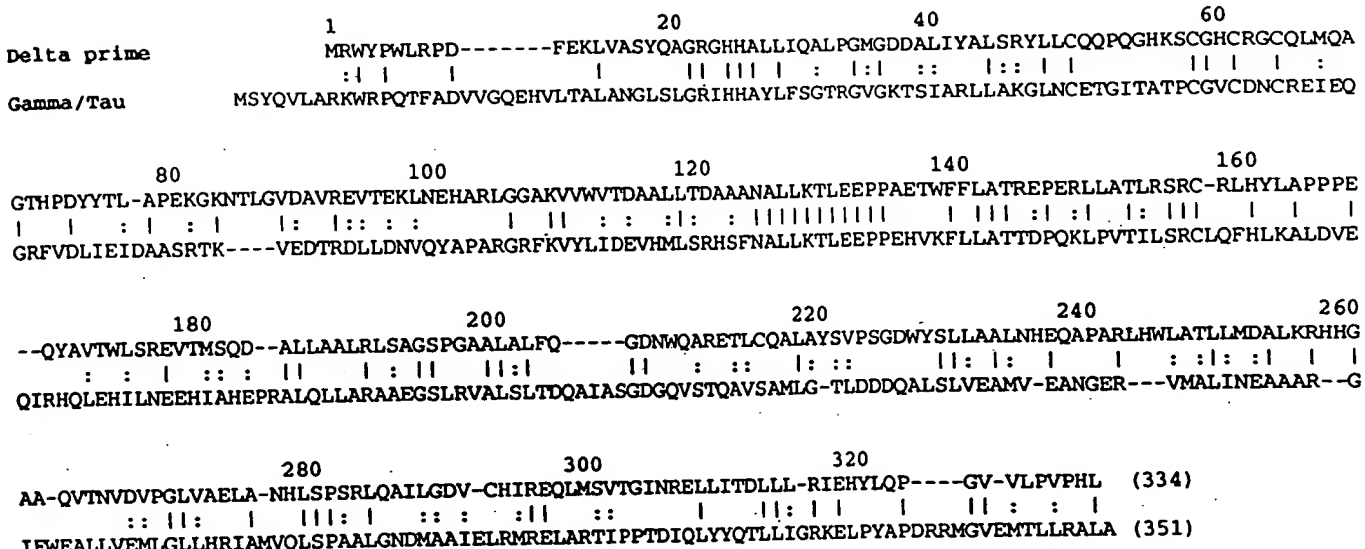


FIG. 9. **Sequence homology between δ' and the γ/τ subunits.** Alignment of the amino acid sequences of δ' and γ/τ (40, 41) subunits of *E. coli* DNA polymerase III holoenzyme. Gaps were introduced to maximize homology. Identical amino acids have a *dash* between them, and conserved amino acids are indicated by the *dots*. Numbering of the sequence of δ' is presented above the sequence. Only the first 351 amino acids of γ/τ are shown (total = 431 for γ and 643 for τ).

chinery serve? It seems possible that the most similar regions of structure among these proteins could serve the same function and that the more dissimilar regions may have different functions. Hence, the coupling of one process to different events may be served by this "protein family" within the holoenzyme particle.

The replicative polymerases of phage T4 and eukaryotes also have multiple accessory proteins (1). The accessory proteins of the eukaryotic polymerase δ (and ϵ) are the 5 protein RF-C complex (or A-1) and the proliferating cell nuclear antigen protein. The accessory proteins of the T4 polymerase are the 44/62 complex and the gene 45 protein.

Recently we have found that the δ' and γ/τ subunits of the *E. coli* γ complex have a significant level of homology to the 36-, 37-, 38-, and 40-kDa subunits of the human RF-C complex and to the gene 44 protein of the phage T4 44/62 complex (38). The homology was especially strong in the 50-amino acid region in which δ' and γ/τ are most homologous. Previously we outlined a homology alignment between β (*E. coli*), proliferating cell nuclear antigen (yeast and human), and gene 45 protein (phage T4) using the structure of β as a guide (10). The homology in both function and sequence among accessory proteins spanning the range from *E. coli* to humans suggests that the basic mechanism of DNA replication has been conserved throughout evolution.

Acknowledgments—We thank Drs. Ken Marians and Kenton Zavitz for helpful advice and Dr. Charles S. McHenry for exchange of the *holA* and *holB* sequences before publication (the sequences match exactly). We are also grateful to Drs. Frank Dean, Mei Chen, and Jerard Hurwitz for the sequences of accessory proteins of mammalian polymerase δ before publication.

REFERENCES

1. Kornberg, A., and Baker, T. (1991) *DNA Replication*, 2nd Ed., pp. 176–181, W. H. Freeman & Co., New York
2. Fay, P. J., Johanson, K. O., McHenry, C. S., and Bambara, R. A. (1981) *J. Biol. Chem.* **256**, 976–983
3. O'Donnell, M., and Kornberg, A. (1985) *J. Biol. Chem.* **260**, 12875–12883
4. Mok, M., and Mariana, K. K. (1987) *J. Biol. Chem.* **260**, 16644–16654
5. Chandler, M., Bird, R. E., and Caro, L. (1975) *J. Mol. Biol.* **94**, 127–131
6. Wickner, S. (1976) *Proc. Natl. Acad. Sci. U. S. A.* **73**, 3511–3515
7. O'Donnell, M. (1987) *J. Biol. Chem.* **262**, 16558–16565
8. Maki, S., and Kornberg, A. (1988) *J. Biol. Chem.* **263**, 6561–6569
9. Stukenberg, P. T., Studwell-Vaughan, P. S., and O'Donnell, M. (1991) *J. Biol. Chem.* **266**, 11328–11334
10. Kong, W.-P., Onrust, R., O'Donnell, M., and Kuriyan, J. (1992) *Cell* **69**, 425–437
11. O'Donnell, M., Kuriyan, J., Kong, X.-P., Stukenberg, P. T., and Onrust, R. (1992) *Mol. Biol. Cell* **3**, 953–957
12. Onrust, R., Stukenberg, P. T., and O'Donnell M. (1991) *J. Biol. Chem.* **266**, 21681–21686
13. Tsuchihashi, Z., and Kornberg, A. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 2516–2520
14. Flower, A. M., and McHenry, C. S. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 3713–3717
15. Blinkowa, A. L., and Walker, J. L. (1990) *Nucleic Acids Res.* **18**, 1725–1729
16. McHenry, C. S. (1982) *J. Biol. Chem.* **257**, 2657–2663
17. Studwell-Vaughan, P. S., and O'Donnell, M. (1991) *J. Biol. Chem.* **266**, 9833–9841
18. O'Donnell, M., and Studwell, P. S. (1990) *J. Biol. Chem.* **265**, 1179–1187
19. Onrust, R., and O'Donnell, M. (1993) *J. Biol. Chem.* **268**, 11766–11772
20. Maki, S., and Kornberg, A. (1988) *J. Biol. Chem.* **263**, 6555–6560
21. Studier, W. F., Rosenberg, A. H., and Dunn, J. J. (1990) *Methods Enzymol.* **185**, 60–89
22. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, pp. 76–85 and 382–387, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
23. Matsudaira, P. (1987) *J. Biol. Chem.* **262**, 10035–10038
24. Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hook. L. E., and Kent, S. B. H. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84**, 6970–6974
25. Marmur, J. (1961) *J. Mol. Biol.* **3**, 208–218
26. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467
27. Kunkel, T. A., Robert, J. P., and Zakour, R. A. (1987) *Methods Enzymol.* **154**, 367–382
28. O'Donnell, M., and Kornberg, A. (1985) *J. Biol. Chem.* **260**, 12884–12889
29. Rowen, L., and Kornberg, A. (1979) *J. Biol. Chem.* **253**, 758–764
30. Pearson, W. R., and Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 2444–2448
31. Konigsberg, W., and Godson, G. N. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80**, 687–691
32. Takase, I., Ishino, F., Waxhi, M., Kamata, H., Doi, M., Asoh, S., Matsuzawa, H., Ohta, T., and Matsuhashi, M. (1987) *J. Bacteriol.* **169**, 5692–5699
33. McHenry, C., and Kornberg, A. (1981) *Enzymes* **14**, 39–50
34. Biswas, S. B., and Kornberg, A. (1984) *J. Biol. Chem.* **259**, 7990–7993
35. Gorbalenya, A. E., and Koonin, E. V. (1989) *Nucleic Acids Res.* **17**, 8413–8440
36. Kohara, Y., Akiyama, K., and Isono, K. (1987) *Cell* **50**, 495–508
37. Deleted in proof
38. O'Donnell, M., Onrust, R., Dean, F. B., Chen, M., and Hurwitz, J. (1993) *Nucleic Acids Res.* **21**, 1–3
39. Edelhoch, H. (1967) *Biochemistry* **6**, 1948–1954
40. Flower, A. M., and McHenry, C. S. (1986) *Nucleic Acids Res.* **14**, 8091–8101
41. Yin, K.-C., Blinkowa, A., and Walker, J. (1986) *Nucleic Acids Res.* **14**, 6541–6549
42. Wickner, W., and Kornberg, A. (1974) *J. Biol. Chem.* **249**, 6244–6249
43. Bradford, M. M. (1976) *Anal. Biochem.* **72**, 248–254

The holA gene encoding δ